Tampere University

Anas Battah

USER STUDY ON USER EXPERIENCE OF SPATIAL AUDIO IN 360 DEGREE MUSIC VIDEOS

Faculty of Information
Technology and Communication
Sciences
Master of Science Thesis
March 2019

ABSTRACT

Anas Battah: A User Study on User Experience of Spatial Audio in 360 Degree Music Videos
Tampere University
Master of Science Thesis, 47 pages, 5 Appendix pages
February 2019
Master's Degree Programme in Information Technology
Major: User Experience
Examiners: Professor Kaisa Väänänen and Dr. Jukka Holm

Keywords: spatial audio, audiovisual experience, thesis, virtual reality, 360 video

With the continuous growth and improvements of visual displays and the high qualities easily accessible by consumers, a need for supporting audio formats grows as well. One of the growing trends over the past few years has been virtual reality and 360° video. Spatial audio provides enhanced perceptions of presence and immersion and thus is crucial to continuous success and expansive VR and 360° applications.

This study focuses on the perceptions of spatial audio in a 360° music video setting, and compares it to the perceptions of stereo audio; both using flat display and head-mounted display. The approach in this thesis is to evaluate four different test conditions with each participant, and compare the results of each participant, as well as between the participants. The four scenarios consist of a music video watched twice on a flat display, once with spatial audio and once with stereo audio at no particular order. Then twice using a head mounted display, once with spatial audio and another with stereo audio in no particular order. The test used evaluation forms with a 7-point Likert scale, in addition to semi-structured interviews.

The interviews aim to gauge music listening habits and the impact they may have on spatial and stereo audio perceptions. The interviews also allow the participants to explicitly state their preference and why, thus providing a better look into the connections between all the different answers.

The results show that spatial audio paired with a head mounted display scored the highest in all our metrics. However, the results from the interviews held after the tests concluded showed less interest in becoming active users of spatial audio. Participants prefer the familiar experience for their day-to-day listening.

The participants predominantly listen to music as a secondary task which works better with stereo audio and is unpleasant with spatial audio. Spatial audio needs to provide a higher value and serve in areas where stereo audio is found lacking. Future studies may focus on that as a topic of research to find the right audience and the right applications for spatial audio.

PREFACE

I am a master's student majoring in User Experience in Tampere University (TUNI), and this document presents my thesis. The thesis is in collaboration with Tampere University of Applied Sciences (TAMK).

TAMK has been for the past 2.5 years working with multiple artists from around Finland to record their concerts in 360° video and spatial audio in order to study the viability of those options and to find the best approach to produce a most impactful experience. With the collaboration of Jukka Holm from TAMK, I have taken the thesis topic of studying the experience of spatial audio in 360° video.

The work of this thesis guided me towards a better understanding of audio formats and the impact they have on a listener, which built upon my background in radio production. As well as opened the path for me to apply methods and practices of research that I learned about throughout my studies towards the degree.

My thanks for the patience, guidance, and feedback of my examiners and thesis supervisors Kaisa Väänänen; and Jukka Holm whose input, previous works, and feedback have been key to the success of this study.

Tampere, 5.2.2019

Anas Battah

CONTENTS

LIST OF SYMBOLS AND ABBREVIATIONS

API	Application Programming Interface
FoV	Field of View
HMD	Head Mounted Display
HOA	Higher Order Ambisonics
SDK	Software Developer Kit
SPS	Spatial PCM Sampling
TAMK	Tampere University of Applied Sciences
TUT	Tampere University of Technology
VR	Virtual Reality
VE	Virtual Environment

.

1. INTRODUCTION

This chapter introduces the topic of the thesis and describes the background and motivation of the study, as well as the research objectives and questions, and finally the structure of thesis as presented in this document.

1.1 Background and Motivation

Virtual reality has seen decades of development and rise and fall stages, dating even further back than digital development [1]. Over the past few years VR has shown unprecedented growth, with the global virtual reality/augmented reality market projected to be 209.2 billion U.S. dollars by 2022, a massive growth from 2016's 6.1 billion U.S. dollars market [2].

Research and development all around the world is working towards finding the best applications for VR and finding ways to access mass consumer markets, while developing better hardware and accompanying software. Many state of the art devices already available allow for display quality as high as 8K such as the Kickstarter crowd funded Pimax [3] (shown in figure 1.1). Another example is the new VR-1 from Varjo, the only VR headset with human-eye resolution display, designed for professionals in complex and demanding industries [4].

Figure 1.1 – Pimax: The world's first 8K VR headset [3]

As VR is an audiovisual experience, the audio delivery needs to match the high level visual displays available in order to achieve the best experience possible for the users. The background of this study comes from Tampere University of Applied Sciences (TAMK) and their work with 360º videos and spatial audio production, more specifically in music and live concerts. One of the productions from TAMK is the Finnish band Popeda's song Helvetin Pitkä Perjantai [5] which was mixed in both spatial audio and stereo audio to give the band the choice, however the band decided to use the stereo audio mix instead possibly due to the technology being too new for the band's conservative audience or possible the band just wanted to keep things simple [6]. The situation paved the way for this study in order to find out which version would actually be the preferred choice for end users.

1.2 Research Objectives and Questions

This thesis is a user study that aims to find the perception of spatial audio and comparing it to that of stereo audio in 360º video using one of TAMK's produced videos in the tests [5] and measure for its appeal to users in live music applications. The research question of this study is: **How does spatial audio perception in 360-degree music videos compare to that of stereo audio perception in 360-degree music videos?** The test had 20 participants, each participant was presented with the music video produced by TAMK with variations in visual display and audio format combinations. The participants were interviewed in addition to answering an evaluation form after each of the variations, the results of the interviews and the answered forms will help in gauging perceptions of spatial audio and compare it to that of stereo audio.

A secondary research question in this study is: **How do listening habits impact perception of spatial audio in 360-degree music videos?** The participants are interviewed in the beginning of the test sessions about their listening habits. The results of the test evaluations compared to the listening habits of the participants will give some insight into possible connections between listening habits and spatial audio perception.

The objective of the study is to *identify patterns in user perceptions* of the different combinations and the impact each of them has on the experience, in order to identify the technology's strengths and pitfalls to help aid future research and development. In addition to *finding out the end users preferences* relating both to the audio formats (stereo audio vs. spatial audio) and displays (flat display vs. head mounted display).

1.3 Structure of the Thesis

Chapter two presents the theoretical background used for this study, the chapter is divided into four sections; the first subsection introduces VR and its presence in entertainment. The second subsection talks about 360º videos and the differences between it and VR,

moving on to spatial audio in the third section, and finally the whole immersive audiovisual experience and theories surrounding that, as this study is directly an immersive audiovisual experience one.

Chapter three delves into the study and the approach used, in addition to the material used, the participants, the variables, metrics, hypothesis, and the whole process of the study. Chapter four presents the results of the study starting the listening habits followed by comparing the different variations, first comparing audio results on each display, followed by comparing each audio format performance on the different displays. Chapter five discusses those results further and finds possible relations between listening habits and the presented results, the limitations of the study, and the conclusions reached in addition to future work and development that could be pursued.

2. THEORETICAL BACKGROUND

In this chapter we take a look at relevant studies and resources that are studied as a part of the thesis. The chapter contains four subsections, talking about different parts of the thesis, starting from introducing Virtual Reality and its role in entertainment, in direct relevance to the focus of the study of music videos.

The following section takes 360º videos into account and compares it to VR. The third section delves into audio, and while focusing on the spatial format; other formats are brought to light and compared. And finally bringing the audio and visuals together and the impact that each has on the other.

2.1 Virtual Reality

This section brings to light different academic and non-academic relevant works. The first subsection introduces VR and the second one delves deeper into VR in entertainment.

2.1.1 Introducing Virtual Reality

The word virtual existed before virtual reality, however its use with virtual reality started because of the virtual images that are viewed through head mounted displays [7]. Virtual reality is produced by simulating scenes as generated through computers to provide the people with the convenience to experience and learn in a virtual world [8].

Another definition used by the early developers to build VR on goes as “A computer-generated three-dimensional landscape in which we would experience an expansion of our physical and sensory powers; leave our bodies and see ourselves from the outside; adopt new identities; apprehend immaterial objects through many senses, including touch; become able to modify the environment through either verbal commands or physical gestures; and see creative thoughts instantly realized without going through the process of having them physically materialized”, according to [5, p. 1].

Taking a step back to virtual reality visions and ideas from 1965 as Sutherland talks about our familiarity with the physical world and its properties, and how a display connected to a computer “gives us a chance to gain familiarity with concepts not realizable in the physical world”, according to [1, p. 1]. Continuing to describe the ultimate display as a “room within which the computer can control the existence of matter” [1, p. 2]. Virtual reality and virtual environments are getting closer and closer to what Sutherland imagined the ultimate display would be.

To get a better grasp on virtual reality, Brooks [11] asks us to think of it as a window, rather than a screen, a window that looks into a virtual world, paraphrasing Sutherland's vision regarding the ultimate display. Brooks separates technologies as crucial and auxiliary for VR; those that are crucial consist of 1) the visual display, 2) the graphics rendering system, 3) the tracking system of the user's head and limbs orientation, and 4) the database construction and maintenance system. Since the 1990s, most of those technologies have come a long way, despite tracking still facing some issues with users reporting nausea and motion sickness from using head mounted displays, though that also relates to the latency of the system.

The important but not so crucial technologies consist of an audio display, including directional and simulated sound fields; other modalities of interaction such as haptic sensations, and other devices that allow for interaction with the VE, allowing for interaction techniques that substitute those possible in the physical world.

In discussion of virtual reality environments, Grigorovici [12] brings three main theoretical statements; 1) the potential of VR environments to become the ultimate mass medium, 2) their association with presence, characterized by high levels of arousal, and 3) associated lower levels of ad awareness.

Despite Grigorovici's statement about VR's potential to become a mass medium, Steuer et al. [13] presents the issue of VR being typically portrayed as such, with a technological focus that has inadequacies failing to provide insight into processes or effects of using the systems, in addition to lack of frameworks and guidelines. However Steuer et al.'s paper is more representative of its time in the 1990s as opposed to a shift of focus nowadays on presence, immersion, and dealing with side effects of VR systems and software alike.

The terms immersion and presence are integral to VR and VE and thus understanding the difference between them is rather important, as compared to using them interchangeably. Slater argues that understanding them separately is required in order to progress VR, reserving "immersion" for what the technology delivers from an objective point of view. Whereas "presence" is the human reaction to immersion, which is subjective. [14]

An example to further explain the difference between immersion and presence, would be listening to a great concert on a high-end audio system and the listener feeling like they are at the concert. Whereas the listener's attention to the content of the music. And so that presence a difference between form, which is relevant to presence and can be induced by the system and its capabilities. Presence is described as just like being somewhere, thus comparing the experience to a more tangible physical one.

Content relates to interest, or attention of a user's, and what draws a users' attention can completely differ between individuals. And as presence is based on perception, and immersion on a system, perception is not dependent on a high quality immersion in order to

take place. Immersion is used in a virtual setting, dealing with a system, however presence and interest apply to day-to-day situations and activities.

2.1.2 Virtual Reality in Entertainment

Virtual reality presents a new medium for art, a medium where the user can be interactive with the content, and have a say in the kind of experience they get, to an extent. VR spans over the spectrum of anything that could be labeled as entertainment, allowing for a variety of options suiting different needs and tastes. A recent example is the incorporation of VR in the NBA (National Basketball Association) to provide the possibility to watch games live in VR, with a courtside experience, thus opening doors for many people to experience what might have seemed out of reach before [15]. The NBA uses Intel True VR to provide those experiences as shown in figure 2.1.

Figure 2.1 – Intel True VR set-up for NBA VR [15]

VR is also used as a way to advocate for different causes delivering a strong message through the technology such as 360labs [16] and their documentary in regards to the grand canyon, among other virtual tours provided part of their services.

Virtual environments provide users with a very high degree of perceptual immersion in comparison with the rest of the media. And as such, their features have significant effects on users' arousal, mood, emotion, and memory. Which makes VR a powerful entertainment medium, where VR-based advertising can have a rather powerful impact, so how does VR perceptual immersion and presence affect persuasion? Entertainment and narrative-based virtual environments 1) could provide a sense of vividness closest to real experiences which have the most powerful impacts on persuasion. As well as 2) having a very strong effect over arousal and affect enhancement [12].

Two rising sectors of VR entertainment worth delving into are games, music, and film, as they are the most heavily consumed forms of entertainment. With the TV and video industry's revenue at 286.17 billion US dollars in 2015, and a projected 324.66 billion US dollars in 2020 [17]. Games hitting a revenue of 108.4 billion US dollars in 2017 [18], and the music industry generating 17.3 billion US dollars in the last year [19], allowing for those industries to be quite lucrative for VR and Augmented Reality (AR) to step in. Which already shows revenue in some aspects, with VR and AR combining to generate 4 billion US dollars of revenue, making it the biggest category under the umbrella of "interactive media" [18].

Dolan et al. [20] discuss the complex relation of VR and 360º video. In these mediums, the storyteller uses cues such as lighting, sound, staging, and others to direct the viewer's gaze, thus tapping into a new realm of possibilities in VR and 360º entertainment applications.

However, the line between movies and games blurs with VR, as there are hybrid forms of film with gaming elements, especially within a virtual environment. The amount of interaction and user input may well decide the categorization. The number of projects per year has been increasing substantially over the years. From two projects in 2014 to 91 in 2017, with the US leading the amount of cataloged content with 60%, and the UK second with 10%. With the VR titles productions primarily being located in the Anglo-American region leads to English being the priority language as it stands now. With the available VR titles, documentaries are particularly popular with them being 33% of the available content, with a common use of 360º cameras. [21] Possibly opening the doors to different markets with accessing different languages, genres, and a diverse expansive user base.

As for the music industry, VR has the potential to revolutionize the way we consume different aspects of music, whether it is music videos, live music, or music education. After YouTube and Facebook launched their 360º video support musicians have taken to posting such formatted and shot videos which can be viewed with a head-mounted display as a VR experience. [22] More on 360º videos is discussed in the next chapter.

A few examples of VR music videos and live music are: Gorillaz – Saturn [23], Popeda – Helvetin Pitkä Perjantai [5], and other worldwide popular bands such as Metallica [24],[25] and Megadeth [26] releasing live 360º recordings of some of their songs, among many other musicians. The list of examples keeps growing thus signalling an increased interest in VR music consumption.

Mbryonic has also developed a platform called Amplify VR where audiences could watch any music video in a reactive immersive VE with the ability to interact with the content. Interactions include the ability to move and remix their own experience, with one of its unique features being its ability to 2D video content to a 3D VR experience [22]. And while VR will most likely not replace completely the thrill and the experience of being

present at an actual concert, it most certainly provides alternatives to those who are unable to be present for any reason. The large and popular music festival Coachella partnered with Vantage.tv in 2016 to provide VR access to both those at the festival and those who couldn't make it, as they made cardboard VR with access to the VR app available for purchase. [22]

With music education, instrument teaching is an obvious aspect to explore with the way the technology is heading. An example of that is Teach U: VR [27] which enables learning or practicing music even without access to an instrument physically. Teach U: VR allows users to play virtual instruments in a virtual environment, drums and piano are incorporated into this project. Another example is Electronauts [28] which is a music creation application that can be experienced in VR.

2.2 360° Videos

360º videos are video recordings that use omnidirectional cameras to capture a space onto a spherical video [29]. The playback of 360º videos the viewer is able to control the viewing direction of the video, with the experiences differing based on the display used. The spherical video captured by the omnidirectional camera is formed by stitching together the various captured perspectives. This is done to generate an immersive experience and an alternate space that places the viewer within the scene rather than presenting it to them as an outside observer and giving them the ability to control orientation and viewing direction [29]. Many options for 360 degree video capture are now available, and websites such as threesixtycameras.com [30] are dedicated to presenting and discussing them.

Some 360 degree video platforms paved the way and have been important players in the field especially pertaining to music videos such as Magenta Musik 360 [31] which is a Dutch website streaming concerts and festivals in 360 degree video. Another company and platform is Jaunt which provided musical content in 360 degree video with artist collaborations to deliver unique content (e.g.: Paul Mccartney) [32], however Jaunt has given up on VR and is shifting their focus to AR experiences from October 2018 [33].

Virtual reality and 360° videos are sometimes used interchangeably, however that is not always the case, as despite some similarities and undeniable synergy, there are some differences to point out. Brooks defines a virtual reality experience as "any in which the user is effectively immersed in a responsive virtual world" [11]. Whereas 360° videos is an enclosed space which a user can view as they wish without interacting with the actual environment, nor does the virtual world respond back. However, there is no reason to discount 360° as a VR experience if viewed in a VE setting.

Multi-camera rigs are utilized to record live action 360º video, giving the consumer a contained perspective to a location. Whereas VR allows for a world in which the user operates as "natural extension of the creator's environment", moving beyond 360º video.

[20] However despite some (such as Dolan) requiring interaction with the content in order to consider it VR, watching a 360 degree video using a HMD may effectively render the experience a virtual reality one as it isolates the viewer from the real world and places them in a virtual one.

Dolan et al. [20] presents different viewing models as follows; 1) The observant model, where the viewer does not have a rigid identity within a story, but merely granted presence through the ability to view the story. Whereas 2) the participant model recognizes the viewer's identity within the universe of the story. And in the 3) active model the viewer is given the ability to affect the outcome of the story's events. Which is an opposite to the 4) passive model within which is the traditional way of storytelling. The first two models define the viewer's existence within the virtual world whereas the latter two models present the interactive influence the viewer has.

As for the worth of going for 360° videos, more specifically in advertisement, Google partnered with Columbia Sportswear to study that. Habig [34] questions what 360° videos can actually do for a brand and whether it ensures higher viewer metrics despite the immersive storytelling that the format promises to deliver. The experiment to find the answers, two similar ad campaigns were created featuring a 60-second spot where one version was shot and presented in 360° video, and the other in standard format video. And to test which format better leads users to respond to answer to the advertisement (e.g. go to an extended version), a call to action button was added to both versions.

After comparing the viewer metrics, the results found that 1) 360° does not over perform with traditional viewer metrics, as users are not always in the mood to interact with 360° video if they're primarily watching standard videos. However, 2) it does motivate viewers to watch more and interact, which came with a lower video retention rate, as viewers did not need to go through the whole cut before wanting to see more. The 360° ad also 3) showed much better results with earned action metrics compared to the standard format ad, such as sharing, channel subscriptions and engagement. As well as increased organic viewer growth for the full-length 360° ad with a 46% higher view count at the end of the experiment, during which both full versions were unlisted, meaning the only way to get to them was through ad-clicks or using the URL directly.

The conclusion from that experiment shows that 360° video has great potential in driving engagement, as it encourages viewers to be a closer part of the action by controlling their perspective, in addition to the novelty of the format making people more interested in both watching those videos and in sharing them. [34]

Despite the experiment being focused on ads, the potential that is shown there is as beneficial in other applications such as music videos, sports highlights videos, or any other relatively short experiences that have the capacity to be shared and spread between users.

With a significantly growing interest towards 360º VR videos, the problem of its extremely demanding bandwidth usage becomes more and more apparent, which makes it more difficult to stream at an acceptable level of quality. Hosseini et al. [35] propose "an adaptive bandwidth-efficient 360º VR video streaming system using a divide and conquer approach.".

The approach is "to deliver higher bitrate content to regions where the user is currently looking and is most likely to look, and delivering lower quality level to the area outside of user's immediate viewport", according to [15, p. 107] , thus focusing on the user's Field of View (FoV) using viewport adaptation techniques. The initial experiments showed up to 72% saved bandwidth on 360º VR video streaming without much noticeable impact on quality. [35]

2.3 Spatial Audio

Spatial audio is "an immersive sphere of audio meant to replicate how humans hear sound in real life" [36]. The following subsections introduce different sound systems followed by discussing different spatial audio recording and playback formats.

2.3.1 Introduction to Sound Systems

With sound systems there are a few terms and definitions that need to be cleared through, as they are most popular, and most relevant to our research. Mono or monophonic describe systems where all audio signals are mixed together and routed through one audio channel. Whereas stereo or stereophonic sound systems have two independent audio signal channels. [37] More commonly known with their numbers, surround sound 5.1 and 7.1 are prime examples of such multichannel sound systems, the numbers referring to the amount of speakers used followed by amount of subwoofer speakers, so five smaller speakers and one subwoofer in 5.1 and seven smaller speakers and one subwoofer in 7.1 with more power and accuracy provided as one goes bigger with the sound systems, however room size and other factors play a role in what setup is best as Boffard describes in [38]. It is possible to go bigger if the financial means are there as it gets more and more expensive with increasing requirements pertaining to room size and others (such as listening position, type of furniture in the room, other preferences), for example a 9.2 setup would have nine speakers and two subwoofers, or another dimension can be included by adding speakers to the ceiling such as the 9.2.4 system [38].

Most commonly in a cinema setting, the Dolby Atmos sound system expands on the previously mentioned surround sound systems. Dolby Atmos uses up to 64 speakers placed around the theatre providing a 3D audio experience, using the height dimension by placing some of the speakers on the ceiling. This creates a hemisphere of speakers allowing sound designers to direct specific sounds to certain areas in the room to a high degree of accuracy. The Atmos technlogy allows for a foundation level of sound mixed

using the traditional channel-based approach, using the static and ambient sounds that do not require specific placements or directions. On top of that layer audio objects are placed along with their spatial metadata in order to create the dynamic sound experience. The technology allows for 128 channels, 10 of which are used for the base layer thus leaving 118 for audio objects. [39]

A simpler than Atmos codec that allows the system to process surround sound is DTS:X which is also the most common as it doesn't require a minimum number of speakers, is purely software based, and has great conversion capabilities [38]. A third highly specialised codec is Auro-3D which relies on a speaker installed in the ceiling; this codec is the least common of the last three mentioned [38].

2.3.2 Ambisonics

Ambisonics is one way to record, mix, and playback spatial audio; in a basic approach, it treats an audio scene as a full sphere of sound coming towards and around a center point, whether it the microphone while recording, or the listener's listening "sweet spot" [40].

The most basic and most widely used Ambisonics audio format is the four-channel B format also known as first-order Ambisonics. First-order Ambisonics uses four channels recorded using four different microphones each pointing in a specific direction while they are all conjoined at the center point of the spatial audio sphere. Within this format, two conventions which are quite similar but not interchangeable are available; AmbiX and FuMa, and they differ by the sequence in which the four channels are arranged. The first order is widely supported nowadays however it is a simple form of Ambisonics. Higher order Ambisonics can provide higher spatial resolutions with the second order utilizing nine channels, the third order using 16 channels, all the way up to sixth order Ambisonics with 49 channels. [40] The Ambisonics orders with channels above four (second order and above) are referred to as higher-order Ambisonics (HOA), and with the higher spatial resolution they provide, accuracy is improved as well [41].

Ambisonics audio and traditional surround sound are sometimes mistakenly confused with one another, however there is a reason Ambisonics were the adopted technology of choice for VR and 360º applications. Ambisonics "can be decoded to any speaker array"; thus representing a full uninterrupted sphere of sound without restrictions of any specific playback system's limitations. Whereas the principle behind traditional surround sound and stereo sound technologies –despite surround sound being more immersive than the latter- go back to the same principle of creating an audio image by sending audio to a pre-determined speakers array. [40]

Ambisonics 1) provide a smooth, stable and continuous sound in a dynamic environment, in contrast to the static environments within which traditional sound formats may prevail. As well as 2) a design that spreads the sound evenly all throughout the sound sphere. And

finally, 3) Ambisonics also provide elevation, where sounds could be represented as coming from above and below in addition to front and behind the listener; in contrast to horizontal dimension limitation of traditional sound formats. [40]

In the end Ambisonics can be played back by decoding the format's channels for the specific speaker arrays, with the result being that resources aligned with the direction of the speaker are louder while ones not aligned are either lower or canceled out. If Ambisonics is played back on a regular stereo setup the entire mix will be folded down to work with the available speakers [40]. Playback is also made possible with the binaural audio technology, through headphones; which "receives an audio input and direction in which to position it." [40]. Binaural audio works in a way similar to our ears which recreates the perception distance. [42]

2.3.3 More Formats and Other Examples

Spatial PCM Sampling (SPS) is a modern alternative to Ambisonics for spatial audio contents such as recording, synthesizing, manipulation, transmittal, and rendering. An SPS multichannel track consists of a bunch of signals recorded by "a set of coincident directive microphones, pointing all around, covering (almost) uniformly the surface of a sphere." Thus SPS signals do not contain time differences between the channels, where only amplitude is different depending on the position of the sound source, and in that SPS finds exact similarity with Ambisonics. SPS -32 records signals simultaneously with 32 "ultradirective virtual microphones" with the use of an Eigenmike. [43]

SPS is found advantageous in most cases when compared to Ambisonics; SPS is much easier to understand, and the signal can be created without complex mathematical formulas. And with a possible large channel count of 32 and more, each sound source could be sent to just one channel thus ridding of the need to "pan" across channels. Panning would still be required for a small number of channels; however that can be done with traditional well known panning functions. The SPS method for rendering the intermediate format to the final loudspeaker system. It also trivializes playback of 360º video with spatial audio soundtracks over VR devices, as it is only necessary to place a spherical distribution of sound sources around the spherical video projection screen, with each being fed with one SPS stream channel. Ambisonics playback on the other hand can get tricky due to the need for an advanced decoder. [43]

Mach1™ is an example application of SPS corresponding to SPS-8; Mach1 is growing as a spatial audio format to use with 360º videos on VR HMDs, ensuring that users with headphones hear a binaural rendering of the spatial scene. [43]

To make spatial audio more consumer facing and increase its accessibility, Nokia introduced OZO Audio which allows for spatial audio capturing using smartphones, including

depth, direction, and detail within one degree of audio accuracy. Using existing phone hardware thus ridding the users of need for extra gear. [44]

Immersive experiences can be created by embedding fitting visual and audio cues into objects in a visual scene, 2D or 3D. Conventional sound systems such as stereo and surround sound are currently used to deliver an audio-visual experience, alongside 2D or 3D display. However, they may not accurately reproduce spatial sound content, such as hearing a non-playing-character getting closer in addition to seeing them come closer. And to achieve this "sound envelopment", surround sound generates the sound around the user; differentiating between left, right, front, and rear speakers. [45]

To overcome the difficulty of accurately reproducing spatial sound using either conventional or directional loudspeakers, Tan et al. [45] proposed and developed a sound system that combines both conventional and parametric loudspeakers, referred to as "the immersive 3D (i3D) sound system". The study concluded that parametric loudspeakers are capable of rendering audio cues from point-like sources, and the ambience effectively reproduced using conventional loudspeakers. The lack of sound overlap, or crosstalk between parametric loudspeakers leads to accurate localization. Thus reaching an improved spatial sound reproduction.

Morrell et al. [46] introduce a music production tool that is based on Ambisonics but does not produce any B-Format signals. The tool breaks from the order structure of Ambisonics and "allows for variable-order and variable-decoder attributes on a per sound source basis" [46, p. 233]. Some of the unique features this tool presents are 1) distance as a user defined parameter that is achieved through gain manipulation. As well as 2) inside panning which places close sound sources inside the loudspeaker array. And 3) reverberation which is produced by transforming the source into B-Format and running it through a plugin to achieve the reverb. This novel approach to Ambisonics gives the composer/sound engineer the control to define the sound field instead of the technology defining it. The composers/sound engineers do not need to worry about designing speaker layouts with this approach.

Spatial audio is now getting increasing support and popularity, and in recognizing the importance of audio on an effectively immersive experience. Huge tech companies are releasing development kits and support for the format, thus encouraging developers to pursue it as well. Those companies include Facebook with their Audio 360 tool allowing users to publish 360º videos on their feed, with spatial audio support with Ambisonics of the first and second order widely in use [36]. HTC Vive is offering a new spatial audio SDK to allow for easier immersive audio development, the SDK supports HOA with very low computing power which is one of its key features [47]. Google VR with a spatial audio rendering engine optimized for mobile VR, which allows users to spatialize sound sources in a 3D space including distance and elevation cues [48].

The Google VR spatial audio API is capable of 1) sound object rendering, which allows the creation of virtual sound sources in a 3D space, and while spatialized, the sources are fed with mono audio data. 2) Ambisonics sound fields, which can be used for background effects and creating a spatial ambience. And finally 3) stereo sounds, which allows the user to "directly play non-spatialized mono or stereo audio files." useful for music and other similar audio. The audio engine supports full 3D first order Ambisonics a spatial audio format. [48]

2.4 Immersive Audiovisual Experiences

In a study revolving around the impact of platform and headphones on 360º video immersion, Tse et al. [49] investigate the industry claim that 360º videos are a powerful tool to create empathy as they are immersive, and that headphones lead to the full immersive experience. For this experiment, two 360º viewing platforms were used, magic window (no head mounted display), and google cardboard (head mounted display); and with and without headphone use.

The study confirmed the prediction; the viewing platform significantly impacts the immersive experience. Thus using google cardboard led to more involvement in the virtual environment, and lower awareness of real surroundings. The use of headphones however improved immersion with the google cardboard, but had an opposite effect with magic window. With google cardboard, the display cuts the user visually from the outside world, and the headphones cut from the sounds of the real world, thus immersing the user more effectively in the virtual environment. [49]

Other notable findings from the study include the suggestion that some genres might be more suitable than others for 360º storytelling, with nature and documentaries being the popular choices between the participants. And that the platform type and use of headphones did not significantly impact every aspect of immersion, as captivation and comprehension remained unaffected. [49]

To evaluate influence of audience noise on different characteristics of presence (immersion, realism, and social presence) in a virtual reality concert experience, Lind et al. [50] recorded a 360 video concert of a local rock band and took recordings of the instruments through the on-stage mixer separately from the audience recordings and put them together in post-production. With concerts being a social experience, and VR not being one just yet, Lind et al. investigated whether audience noise would affect that.

While auditory feedback in 360 video experiences is usually conveyed with headphones and a head mounted display, Lind et al. chose a high fidelity auditory display in the form of a 64 channels Wavefield synthesis system (WFS), while still using Samsung Gear VR for visual display, a low fidelity display. In the experiment, audience noise showed no significant impact on any presence component.

The fidelity distance between the auditory and visual displays however produced interesting results, as it led to a strong negative audio-visual interaction, the low quality visual display led to perceptions of the experience to be of bad quality. Thus the study found that a low quality visual display reduced quality perception of a high quality auditory display. Which was confirmed by removing the head mounted display and placing a blindfold on the participants while listening to the concert using the same auditory display system. Participants reported a high sense of presence and a higher experience quality as a whole. [50]

In another study, Storms et al. [51] argues that a problem lies in the common consideration that the realism of virtual environments is a function of visual and auditory fidelity mutually exclusive of each other. The problem being that the user of the virtual environment is human, a being multimodal by nature. And as such, the fidelity requirements of virtual environments also needs to be based on multimodal criteria comprising all of the human senses.

With the approach of an experimental psychologist, a series of three experiments took place to investigate the existence of audio-visual cross modal perception interactions. With two independent variables being visual and auditory display quality each consisting of low, medium, and high qualities. The effort aims to answer the question "in an auditory-visual display, what effect (if any) does auditory quality have on the perception of visual quality and vice versa?" [29, p. 558]

The first experiment was on static resolution, which "investigates the perceptual effects from manipulating visual display pixel resolution and auditory display sampling frequency" [29, p. 562-563]. The experiment's findings suggest that when manipulating visual display pixel resolution and auditory display sampling frequency 1) an increase in perception of visual display quality is caused by a high-quality visual display coupled with high quality auditory display when attending to only visual modality or both auditory and visual modalities. 2) When the focus modality is auditory only or both auditory and visual, a low-quality auditory display and a high-quality visual display cause a decrease in auditory display quality perception. And 3) a high-quality auditory display coupled with low-quality visual display causes an increase in auditory display quality perception when attending to both auditory and visual modalities.

In the second experiment with static noise, Storms et al. investigate the perceived effects from manipulating Gaussian noise levels in visual and auditory displays where the visual display consists of a static image of a radio coupled with a selection of music for the auditory display. The findings suggest that 1) a low-quality auditory display coupled with a high-quality visual display causes a decrease in perceived audio quality when attending only to the auditory modality. 2) While attending to only the auditory modality or both auditory and visual modalities, an increase in perceived visual quality is caused by a coupling of high-quality visual and auditory displays. And 3) with the coupling of medium-

quality auditory and visual displays while attending to both auditory and visual modalities an increase in perceived auditory quality is noticed.

The two experiments used a coupling of radio and music as visual and auditory displays. For the third and final experiment, auditory and visual displays that are not semantically associated with one another are used in order to test whether the findings from the first two experiments would hold true nonetheless. The static resolution non-alphanumeric experiment is "designed to investigate the perceptual effects from manipulating visual-display pixel resolution and auditory display sampling frequency." [29, p. 275].

The findings from the last experiment suggest that when manipulating both visual display pixel resolution and auditory display sampling frequency 1) an increase in perceived visual quality is noticed when attending only to the visual modality using a high-quality visual display and a medium-quality auditory display. While 2) an increase in the perception of visual quality is caused by the coupling of high-quality auditory and visual display when attending only to the visual modality, or to both auditory and visual modalities. However 3) attending to both modalities with a medium-quality auditory display coupled with low-quality visual display caused a decrease in perceived audio quality.

The results of those experiments provide empirical evidence that supports previous suspicions across industries; auditory displays can influence quality perception of visual displays, and vice versa. [51]

On spatial audio production for 360 degree live music videos Holm et al. [6] discusses the different aspects of audio mixing for such multi-camera productions. The production work flows were developed and fine-tuned through multiple case studies across different music genres to test whether the production tools and techniques are equally efficient for mixing different types of music. Holm et al. used the Nokia OZO camera in all their video capture projects related to their study; one of the videos recorded and mixed is the Finnish band Popeda's Helvetin Pitkä Perjantai [5] used for the thesis work. Despite the spatial audio mix provided to the band they decided to stick to what is familiar and used the stereo audio mix. The paper concludes with the need for adaptability with the changing and developing nature of spatial audio technologies and speaks about the importance of understanding techniques ahead of what the 360 degree video players such as YouTube are capable of (first-order Ambisonics) [6].

Chang et al. argue that first and second order Ambisonics "are not enough to accurately reproduce sound at ear positions" [52, p. 341]. Chang et al. analyse the impairments/artefacts of binaural reproduction in spectrum and sound localization with three different virtual loudspeaker layouts. The different layouts are to inspect the impact of the layout on the impairments, if any. The results of the study show that impairment occurs when using more than four virtual loudspeakers, which is the number of components of first-

order Ambisonics. The study concludes that localization performance can only be improved by using higher orders of Ambisonics. [52]

3. METHODS AND MATERIAL

This user study brings together a mix of qualitative and quantitative data gathering methods, in order to answer the questions "How does spatial audio perception in 360-degree music videos compare to that of stereo audio perception in 360-degree music videos?" and "How do listening habits impact perception of spatial audio in 360-degree music videos?". And to find out the worth of spatial audio for end-user, and in turn find out some of the value for content creators and artists, to create for spatial audio. This chapter shows the approach, processes, and methodologies used for this study.

3.1 Research Approach and Process

In order to find answers to the questions asked, the test included quantitative evaluation forms and background questionnaires to understand listening habits and first impressions from the scenarios view. A scenario in this test refers to the combination of visual display and audio format used, with two different visual displays and two different audio formats bringing the total number of scenarios to four. In addition to semi-structured interviews to get a better understanding of the participants and relating the potential impact their pre-existing habits have on their experience.

The four experimental scenarios were all presented to all participants with the flat display variations (2D video) presented first, followed by the head mounted display (3D video), with the audio variations randomised in order between different participants and within each participant's experiment (which is first, spatial or stereo), without telling the participants which audio is coming next to test whether participants are able to distinguish the different audio scenarios by themselves. The scenarios are further referred to as related to their combination with PC referring to flat display scenarios and VR to head mounted display scenarios, and stereo and spatial refer to the audio format used, and the scenarios are then as follows; **PC – Stereo**, **PC – Spatial**, **VR – Stereo**, and **VR – Spatial**.

The experiments took place in a room with no external sources of noise that could interfere in the experience, in addition to the use of a pair of headphones with the active noise cancelation feature.

Participants were taken one at a time without contact with other participants on different days over a period of 4 weeks, with each experiment lasting under an hour from start to finish. Participants were led to the room where the experiment took place and were asked to sign a consent form to allow the audio recording of the experiments, which was followed by an explanation of the experiment and what is expected of them to do. Afterwards, each participant was presented the background information questionnaire. An in-

terview was held for each participant, and then once ready the scenario viewing commenced. After each scenario, the floor was open for comments and questions, in addition to an evaluation form to give feedback on the last viewed scenario.

Once all scenarios have been viewed, an interview was held to get qualitative information on the participant's thoughts, feedback, and suggestions relating to the different scenarios.

3.2 Material

A 360 video from a concert for the song Helvetin Pitkä Perjantai [5] by the Finnish band Popeda with two different sound editing variations, one produced using stereo mode, and the other produced in 3D/spatial audio mode using 1st order Ambisonics.

The two variations were then presented using different displays, the first being a flat screen display, and the second being a head-mounted display (Samsung Gear VR) used with Samsung Galaxy 7 Edge, with Samsung Galaxy 7 as back-up. With all audio being heard through the same headset (Bose QuietComfort 35 Series I), providing consistency in the highest quality possibly achieved.

The study uses a headset for all scenarios due to the nature of spatial audio and that it would be rendered ineffective with the use of loud speakers. Headsets were also used in the stereo audio scenarios in order to maintain consistency across the test.

3.3 Sample

The sample consisted of 20 participants (15 male and five female), gender based differences were not a focus of the study, however are taken into account in the analysis of the results. With ages ranging from 22 and 34 years old (Mean = 26.10). Participants knew about the study and took part in it mostly through word of mouth and referrals from colleagues and acquaintances, and all went through the same experiment process.

Out of the 20 participants, 9 were hobby instrumentalists with a range of different instruments, instruments played is irrelevant to the test. However playing an instrument is assumed to have an effect on perceived audio quality and attentiveness to instruments played in the test video. The participants also answered questions on a 7-point Likert scale to determine their familiarity with different technologies used in the test namely their familiarity with VR, 360 degree videos, 360 degree music videos, and spatial audio, with median scores of 3.0, 3.0, 1.0, and 2.0 respectively. The scores are rather low signalling generally low familiarity with the technologies, with many being introduced to those technologies for the first time in the test.

With VR familiarity five participants (25%) are completely unfamiliar with VR with a score of one, while 80% of the participants gave a score of four or below. With a slightly

higher familiarity scores, 360 degree video familiarity has only three participants (15%) completely unfamiliar with a score of one, while 75% of the participants gave a score of four or lower. 360 degree music videos results show least familiarity with 11 participants (55%) completely unfamiliar with them with 90% of the participants giving a score of 3 or lower, with the two remaining participants giving scores of six and seven. Despite less participants being completely unfamiliar with spatial audio at nine participants (45%), the general familiarity levels are rather close to the prior technology with 90% giving a score of four or lower.

While the music video used in this test is in Finnish, not all the participants spoke the language or were previously familiar with the artist, however participants from Finland knew the band and had varying opinions and feelings towards the artist, though the impact those factors have on the experience are not a part of this study.

All participants experienced the four variations of the material, however in a randomised order, with flat display variations always coming first.

3.4 Variables

The independent variables are SOUND (stereo sound and spatial sound), DISPLAY (flat screen and head-mounted display), GENDER (male and female), and INSTRUMENT_SKILLS (hobbyist and no instruments).

Dependent variables are perceived audio quality, perceived stage presence, pleasantness of music and overall experience, and the effect that the choice of music has on the experience regardless of it being positive or negative.

3.5 Metrics and Methods

Two metrics and two interviews were used in this study, a **demographic background questionnaire** presented at the beginning of the test session, and a **user evaluation form** that uses a 7-point Likert scale presented after each video to determine perceived presence, quality, and overall experience subjectively for each user, for each of the presented variations. Both interviews are semi-structured, the **first interview** is held before the videos are presented designed to help better understand the music listening habits of each participant, and the **second interview** to discuss the scenarios and the participant's preferences once the scenarios have all been viewed.

With the metrics and methods provided, we were able to collect both quantitative background data with the background questionnaire (such as age, gender, education, previous familiarity with different aspects of the experiment such as VR, spatial audio, and 360 video, and the ability to play musical instruments), as well as qualitative data from the

interviews. The forms and interview questions can be found in the Appendix at the end of this document.

3.6 Hypothesis

The hypothesis is that **users are most likely to prefer spatial audio within a VR experience in comparison to other variations presented in this study**, due to heightened stage presence from the user's choice of where to focus their attention, and the audio focus changing accordingly. It is hypothesized also that **background information such as gender and education would not have an effect on the prevailing preferred variation**. Furthermore, it is hypothesized that **listening habits would have an effect on preferred variation out of the four**.

4. RESULTS

This chapter shows the findings of the test, the sections are divided into three main sections, the first delving into the listening habits of the participants. The second section discusses the results from the individual test scenarios. A scenario is -as described earlier in this document- the combination of visual display and audio format used, with two different visual displays and two different audio formats.

A 7-point Likert scale was used in the video evaluation forms after each of the video scenarios was presented to a participant, the exact phrasing of the questions can be found in Appendix B.

The effect that the choice of music has on the overall experience only differed slightly between scenarios for each user if any at all. The mean of the means from different scenarios is 4.87 which indicates a slight impact, regardless if it is negative or positive. While that may not be a significant result, it is an indicator that providing choice for users and allowing them to use the technologies according to their preferences may have a growing impact on those technologies.

Scenario	PC - Stereo	PC - Spatial	VR - Stereo	VR - Spatial
Mean	4.60	4.80	4.97	5.12

Table 4.1 – Music choice impact on experience in each scenario (on the Likert scale)

From table 4.1 we can see that the widest difference in means is a mere 0.525, between the VR – Spatial scenario, and the PC – Stereo scenario. And despite it being a marginal difference in the mean between them, there is a seemingly different impact a scenario has on the level of impact a music choice can have on the experience, with the least being on a display screen. Adding the 3D or spatial effect to the audio adds to the experience and the music choice impact, as it shows an increase in the mean in both PC display, and in VR display. And VR as a display shows higher means as a display as well, in comparison to PC display.

4.1 Listening habits

With listening habits, the results shown are the ones that were claimed dominant by each user in the interviews, as most answers are situation dependent and differ from time to time, with a dominant behaviour visible. That is the behaviour that is documented for this study as deemed most relevant.

Even with varying levels of care about the audio quality, none of the participants considered themselves a Hi-Fi listener, some expressed their wish to become as such once it is within their means.

From table 4.2 we see that most of our test participants mainly listen to music as a secondary or background task as long as it does not interfere with the main task. Main tasks included being on a commute, doing sports, studying, or working. Main tasks differed slightly between participants according to personal preference, most notably combining music with a focus intensive task such as studying, compared to house chores such as cleaning or cooking.

Dedicated listening consists of putting the time to listen to music as the main task, allowing for the music to take hold of the moment. This way of listening may have seemed to be a vanishing habit, however it has become a niche in the recent years, especially with the comeback of LP records [53]. LP records are gaining more traction as those who do dedicated listening savour the music as its own experience. Those people are usually either heading towards Hi-Fi systems, or are already using such systems.

As we see in table 4.2 the amount of participants that dominantly listen to music as a main task are a mere 10% of the participants, with 25% depending on mood and situation, and the remaining 65% listening to music in the background. While this could be an indication towards the listening habits of mass consumers, more tests could be done that focus on the different types of listeners, with some focused on users who are more focused on dedicated listening, such a study could provide insight towards the ease of transition towards spatial audio in 360 degree music videos. The other study could focus on people mainly listening to music as a secondary task to gain insight on targeting factors that could be most successful in attracting them towards spatial audio and 360 degree music videos and a more dedicated listening experience.

	Number of Participants	Percentage
Background Listening	13	65.0%
Dedicated Listening	2	10.0%
Mood Dependent	5	25.0%

Table 4.2 – Listening habits

As for listening setups, the test showed that the situation, environment, and timing of listening to music have a large impact on the chosen setup to listen to music, as being at work with colleagues would for example dictate using a headset, similar to being in public

transport. Whereas being at home with friends would render headsets useless, and loudspeakers would need to be used. The reliance on situation and mood shows a substantial percentage within the test of 45% of participants not having a dominant or preferred listening setup (as shown in table 4.3), whereas 45% prefer headsets or dominantly incorporate them for their listening, and only 10% that prefer or dominantly use loudspeakers. And thus at least 90% of the participants would be used to the use of headsets and such a switch to spatial audio would not further require a change of listening setup for them, as can be experienced with what is available to them already.

	Number of Participants	Percentage
Headsets	9	45.0%
Loud speakers	2	10.0%
Situation Dependent	9	45.0%

Table 4.3 – Listening setup

When asked about their dominant behavior when listening to music as an audio only experience or as an audiovisual experience, none of the participants expressed preference in audiovisual experience when it comes to music. Table 4.4 represents the preference results towards an audio only listening experience or watching a video accompanying the music (audiovisual experience). However when asked about live concerts as an audiovisual experience, participants were found to rethink their answer leading them to say that live concerts are a different case scenario especially accompanied with VR.

The natural inclination of the participants was to think of audiovisual experiences with music as watching a video clip with or of the song itself. Such a dominant behavior of an audio only experience does not mean exclusivity. As an example, most participants expressed that they would watch a video clip if it was recommended by a friend or even just merely out of curiosity.

	Number of Participants	Percentage
Audio	18	90.0%
Mood Dependent	2	10.0%

Table 4.4 – Participants listening to music with audio only vs with video

4.2 Audio Format Comparisons: Stereo Audio vs. Spatial Audio

This subsection presents and compares the results of stereo audio and spatial audio tests on each of the displays used in this study (flat display and head mounted display).

The minimum and maximum values in the tables refer to lowest and highest evaluations given for each metric in the title scenario. Despite the content being the same throughout all the scenarios, the delivery is different thus leading to different results from the participants.

4.2.1 Flat Display Scenarios

Flat displays are the most common way to consume audiovisual content, despite the type or the kind of display used. In this study, a computer/laptop screen is used for visual display connected to a mouse for interacting with the 360º nature of the video. Flat display is used to introduce spatial audio to the participants within the context of the study.

The metric "Music pleasantness" refers to the subjective pleasant feeling the user gets listening to it. The table 4.9 below shows the reported minimum, maximum, mean, and median values from the participants in regards to how pleasant the music was. While the difference may not be a large one between the different audio formats when it comes to music pleasantness, it still could be a weak signal that spatial audio is more pleasant than stereo audio.

While the minimum reported value of one in this metric is the only one it is not an outlier as shown in the boxplot in figure 4.1. However, the same participant reported a music pleasantness of six in spatial audio, and the great difference in value may be attributed to technical issues occurring during the test. The participant reported sound buzzing and a *"not so great"* quality while listening to stereo.

	Minimum	Maximum	Mean	Median
Stereo audio	1,00	7,00	4,35	5,00
Spatial audio	2,00	7,00	4,92	5,00

Table 4.5 – Music pleasantness in audio formats paired with flat display (on the Likert scale)

The boxplot also shows a wide spread of opinions regarding music pleasantness in stereo audio whereas in spatial audio opinions are comparatively closer to one another despite

the medians being the same. The main difference comes from the 50% scores of the participants in the middle with a wider variations in opinions shifting towards lower scores in stereo audio compared to spatial.

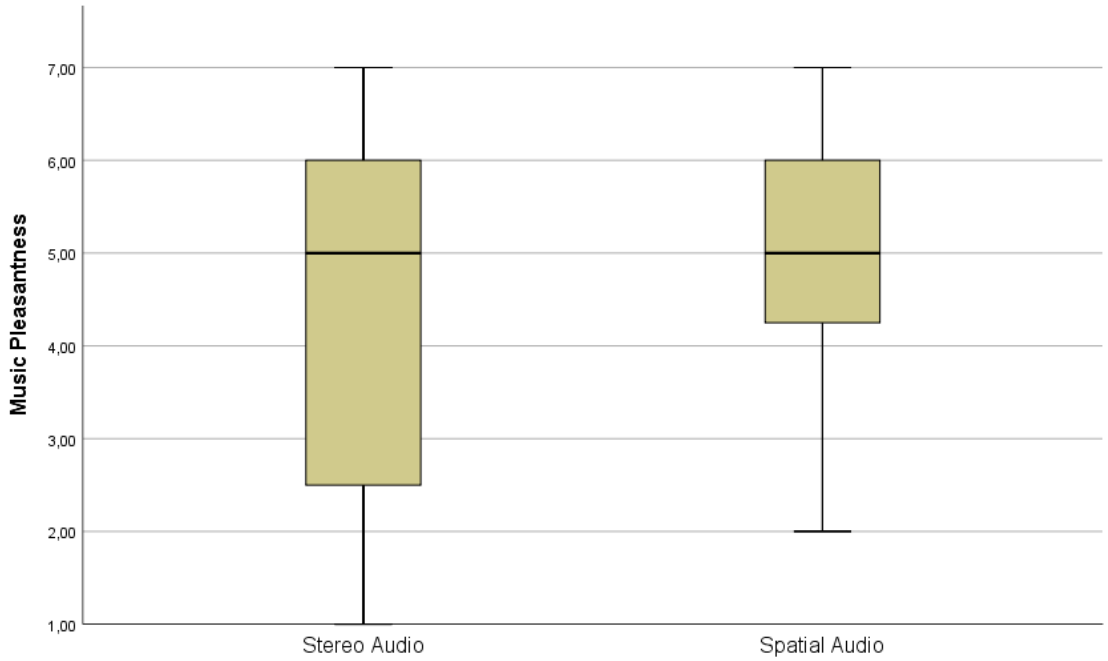

Figure 4.1 – Boxplot of music pleasantness in audio formats paired with flat display

The difference in the mean of perceived audio quality is a small one. Four participants valued the audio quality perceived below three in stereo audio, while three was the minimum from all participants in spatial audio. This potentially gives an advantage for spatial audio over stereo despite the difference in mean for this metric being smaller than that in music pleasantness.

	Minimum	Maximum	Mean	Median
Stereo audio	1,00	7,00	4,70	5,00
Spatial audio	3,00	7,00	5,05	5,00

Table 4.6 – Perceived audio quality in audio formats paired with flat display (on the Likert scale)

Despite the spread of responses from participants in the top 75% being similar between stereo and spatial audio in perceived audio quality (as shown in figure 4.2), spatial audio shows an advantage in the lower 25% scores. The lower scores in stereo audio could be due to technical errors during the test. Such as the headset not being properly plugged which may not come across as clearly as a problem in stereo audio but can impact the spatial audio experience greatly.

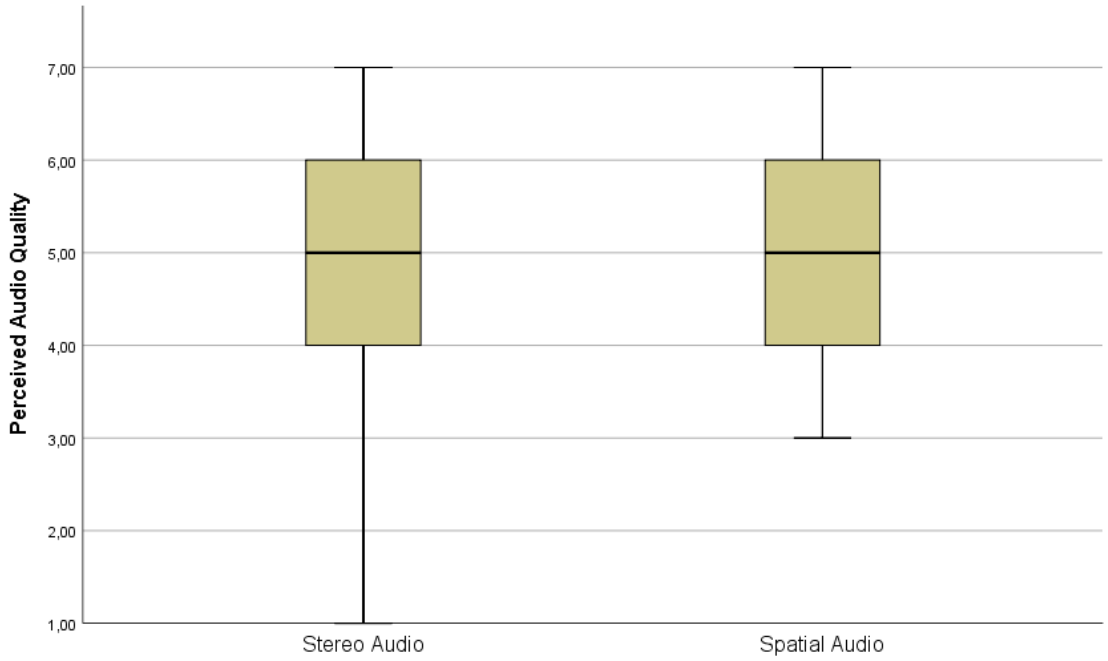

Figure 4.2 – Boxplot of perceived audio quality in audio formats paired with flat display

Perceived stage presence has a higher mean in spatial audio than it does in stereo audio, which is an expected outcome due to the nature of spatial audio that aims at increased immersion. Three participants however reported values of perceived presence higher in stereo audio compared to that of spatial, but the dramatic increase in values from other participants going from stereo to spatial offset the overall mean towards higher immersion when using spatial audio.

	Minimum	Maximum	Mean	Median
Stereo audio	1,00	7,00	3,82	4,00
Spatial audio	1,00	7,00	4,29	5,00

Table 4.7 – Perceived stage presence in audio formats paired with flat display (on the Likert scale)

Figure 4.3 indicates that 75% of participants gave perceived stage presence a score of four or higher in spatial audio compared to 50% in stereo audio, thus showing an increase in stage presence in at least 25% of the participants, while the low scores may be attributed to the visual display used, results from HMD tests could prove or debunk that theory.

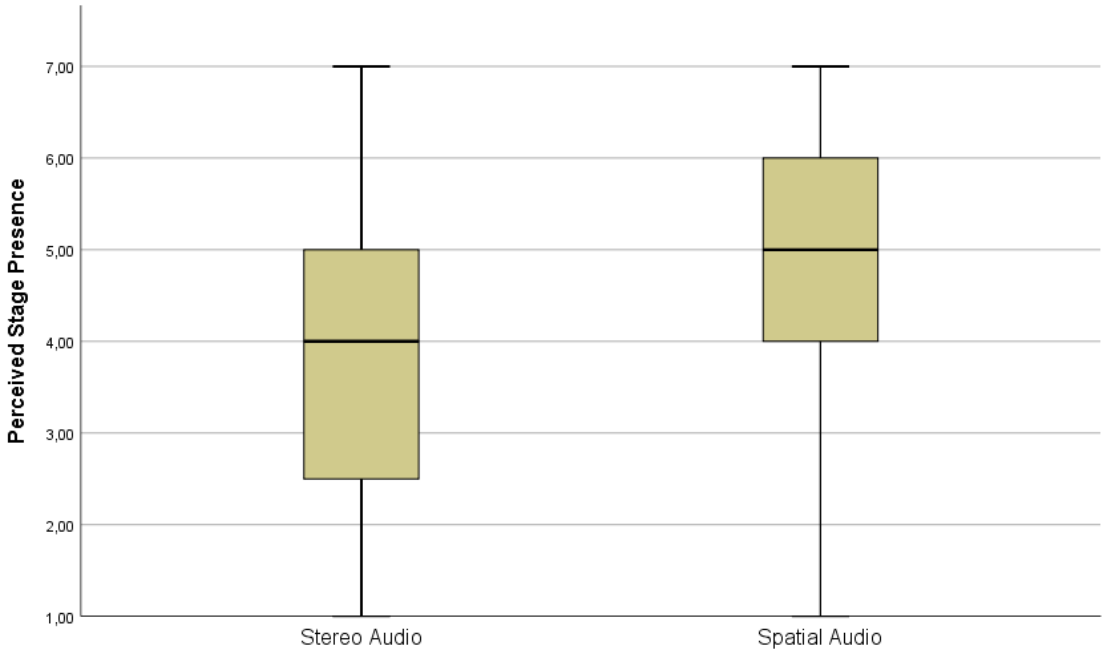

Figure 4.3 – Boxplot of perceived stage presence in audio formats paired with flat display

Table 4.12 shows that the overall listening experience of spatial audio also scores a higher mean than stereo audio among the participants, which could be an outcome of the results from the previous metrics as they are all factors that affect the overall experience.

	Minimum	Maximum	Mean	Median
Stereo audio	1,00	7,00	4,20	4,50
Spatial audio	1,00	7,00	5,07	5,50

Table 4.8 – Overall listening experience in audio formats paired with flat display (on the Likert scale)

The difference in means is also reflected in the difference in medians as well as distribution of scores (as shown in figure 4.4) with 75% of the participants giving a score of four or above in spatial audio compared to the three or above scores registered by the same percentage of participants in stereo audio, the difference present in the distribution of given scores is small and may be insignificant.

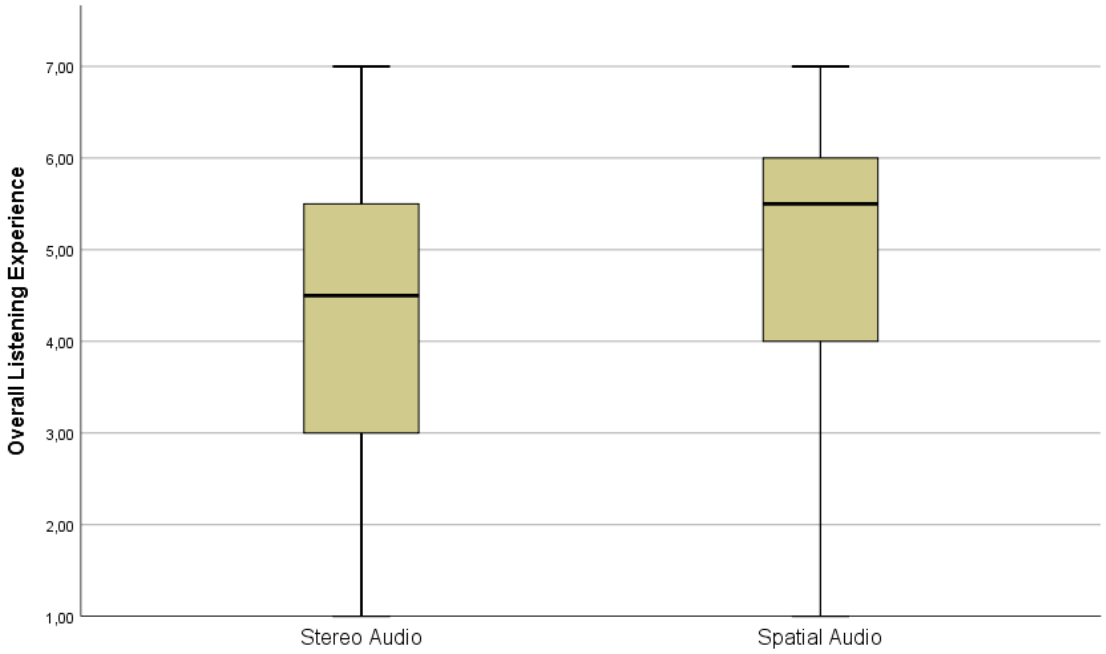

Figure 4.4 – Boxplot of overall listening experience of audio formats paired with flat display

While not a majority of participants increased their evaluations from stereo to spatial compared to those who decreased it, the increased values make considerable jumps as far as going from an evaluation of one to six from one participant. With only a decrease of only one or two evaluation points where spatial audio is deemed the lesser format at that metric, which is one factor as to how spatial audio scores a higher mean than stereo audio.

Spatial audio scores higher in all of these metrics on flat display, which might be a signal to it being a possible preference as a listening experience. However these differences are rather small and almost negligible as the changes of value notes between stereo and spatial did not differ greatly for each of the participants. Three participants (15%) reported that they did not notice a difference at all between the audio formats, despite recording slightly different values on their evaluation forms. Thus the results regarding the audios remain inconclusive.

The figures 4.2 and 4.3 present stereo audio and spatial audio results respectively in a comparison between hobbyist and non-instrumentalist participants provide a curious look at different perceptions of both types audio. In both audio formats, hobbyist participants had higher means in presented metrics in varying degrees from non-instrumentalist means. While the differences in means are definitely there and provide a curious possible point, the difference in perceptions between hobbyists and non-instrumentalists is not a focus of this study and thus those are results are also inconclusive and cannot confirm or prove any theories regarding the matter. In perceived audio quality, the difference between hobbyists and non-instrumentalists is reduced as the latter perceived higher quality in spatial while hobbyists perceived lower quality, the result could be attributed to a more trained ear from hobbyists towards instruments that allows them to potentially notice

flaws as the sound focuses on instruments according to their listening orientation, whereas non-instrumentalists are hearing the clear instruments (in the cases of the participants who noticed a difference between the audio formats) which could lead to a higher audio quality perception.

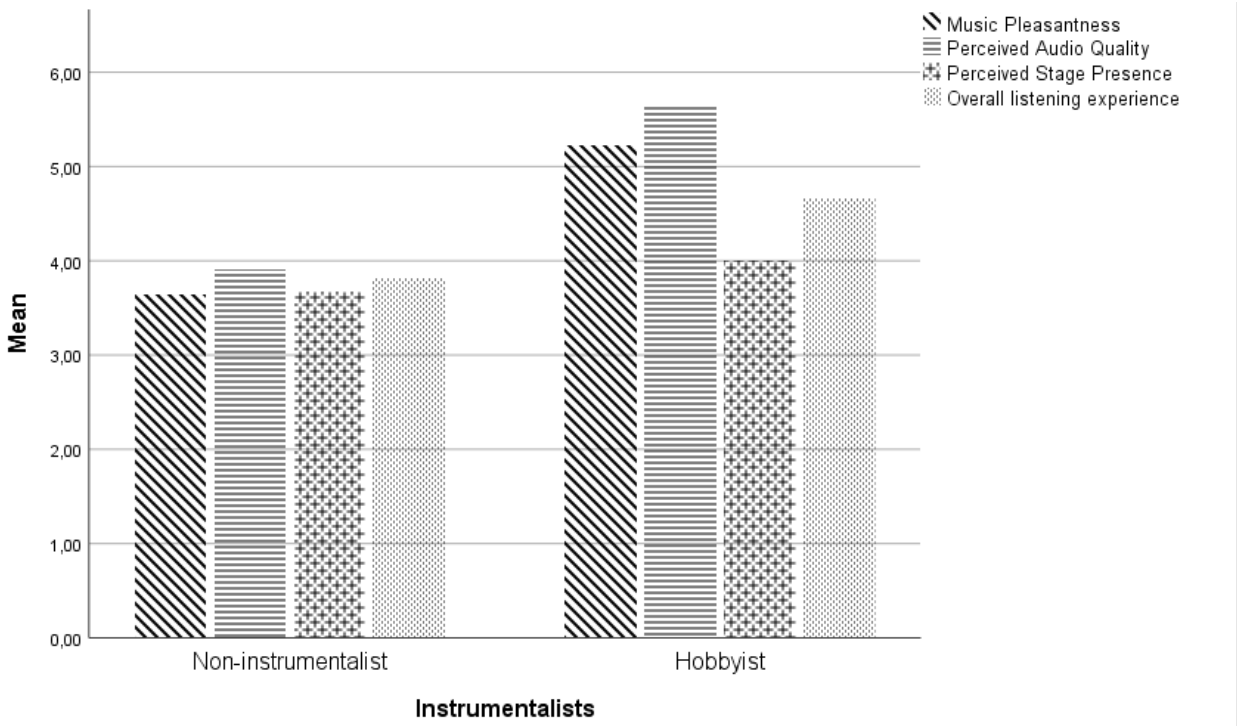

Figure 4.5 – Hobbyist and non-instrumentalist results means for stereo audio

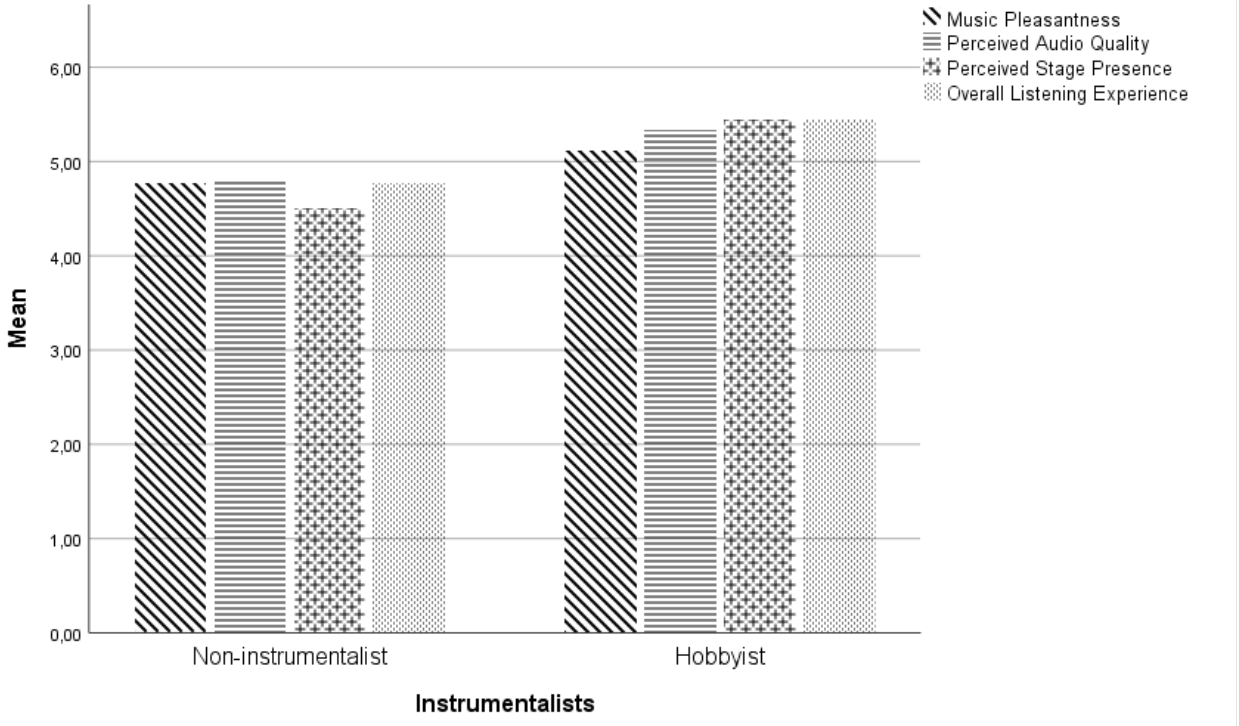

Figure 4.6 – Hobbyist and non-instrumentalist results means for spatial audio

4.2.2 Head mounted display scenarios

Head mounted display is the main way for virtual reality consumption. HMD also provides more immersive medium for viewing and interacting with 360º videos compared to using flat displays. The minimum, maximum, and mean from the different metrics are rather close between stereo audio and spatial audio in the head mounted display tests. A possible factor in that is the novelty of VR and the "wow factor" that most participants expressed. This lead to a focus on the VR experience rather than the audio differences between the two variations. One participant commented that it *"somehow doesn't matter what you are listening to"* as VR provided them with a whole experience, the same participant commented that VR gives an *"illusion of better quality"*.

Table 4.13 shows the values for music pleasantness, which skew more towards stereo audio in their mean. The minimum value recorded is similar, however stereo audio has only one recorded 2 value for pleasantness while spatial has two responses of the same value. The next value recorded for spatial is four, opposed to a three in stereo before going to values of four. The slight difference of 0.2500 in means may be due to spatial audio recording only four values at 7 compared to six responses at maximum value in stereo, giving it the edge. A possible factor impacting this result is the chance at putting the headset on the wrong way around (left speaker of the headset on the right ear and vice versa). Such an error has a very strong impact on the experience in spatial audio while almost none in stereo. A participant who has a trained ear due to playing some instruments reported that the sounds were coming from the wrong part of the stage, which was only attributed to possible headset misplacement after the test concluded.

	Minimum	Maximum	Mean	Median
Stereo audio	2,00	7,00	5,60	6,00
Spatial audio	2,00	7,00	5,35	6,00

Table 4.9 – Music pleasantness in audio formats paired with HMD (on the Likert scale)

While the median, minimum, and maximum values are the same between the two audio formats in HMD, the distribution of the values tells a different story (as indicated in figure 4.7). The minimum scores in spatial audio come from two outliers, the first one coming from a participant who was observed skipping the videos on different occasions, when asked about it the participant mentioned that it is because *"I don't like this of music"* indicating a very big impact caused by the music choice, leading to the participant not wanting to go through the whole experience but only parts of it in order to finish the test and be able to answer the evaluation forms, the participant also reported not hearing a lot of difference between the audio formats and requested to hear them briefly again, upon which they perceived spatial audio to have a lower quality.

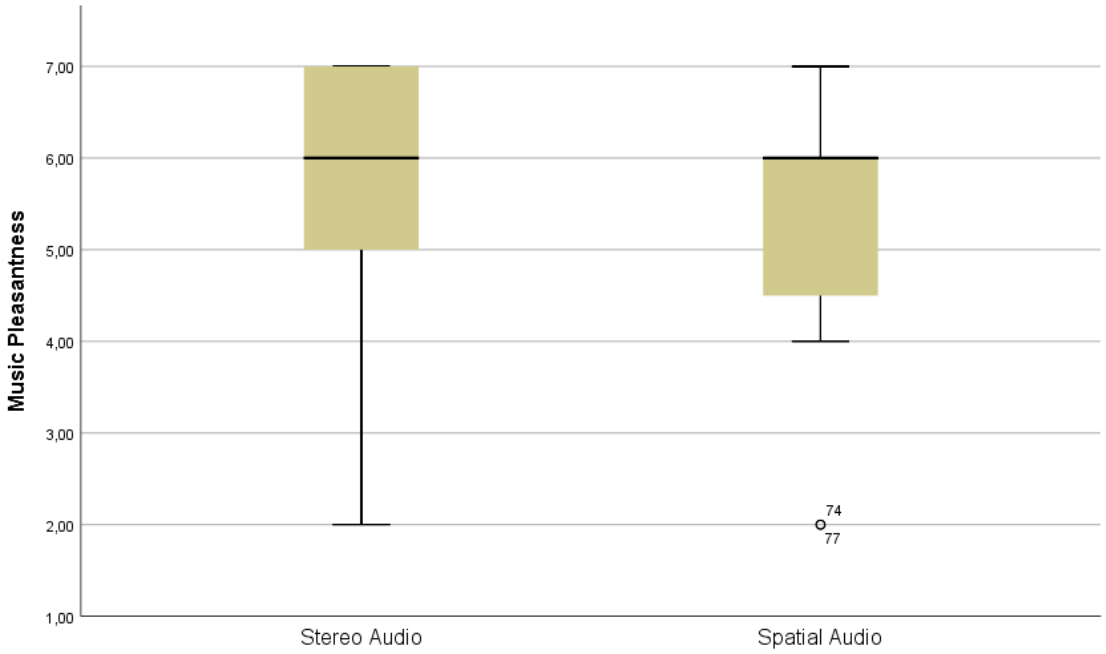

Figure 4.7 – Boxplot of music pleasantness in audio formats paired with HMD

Perceived audio quality has the same mean in both stereo and spatial audio –as shown in table 4.14- despite the individual answers being different and participants perceiving it at different levels. That could signal that the change is not an objective one towards either format and that subjectively either one could score higher.

	Minimum	Maximum	Mean	Median
Stereo audio	2,00	7,00	5,40	5,50
Spatial audio	2,00	7,00	5,40	6,00

Table 4.10 – Perceived audio quality in audio formats paired with HMD (on the Likert scale)

The perceived audio quality is close with all measures despite the distribution of the scores being slightly different with the minimum of two coming from an outlier in stereo audio as indicated by the boxplot in figure 4.8. The outlier comes from the same participant who kept skipping on the videos however it does not have a large impact on the results or the conclusions drawn.

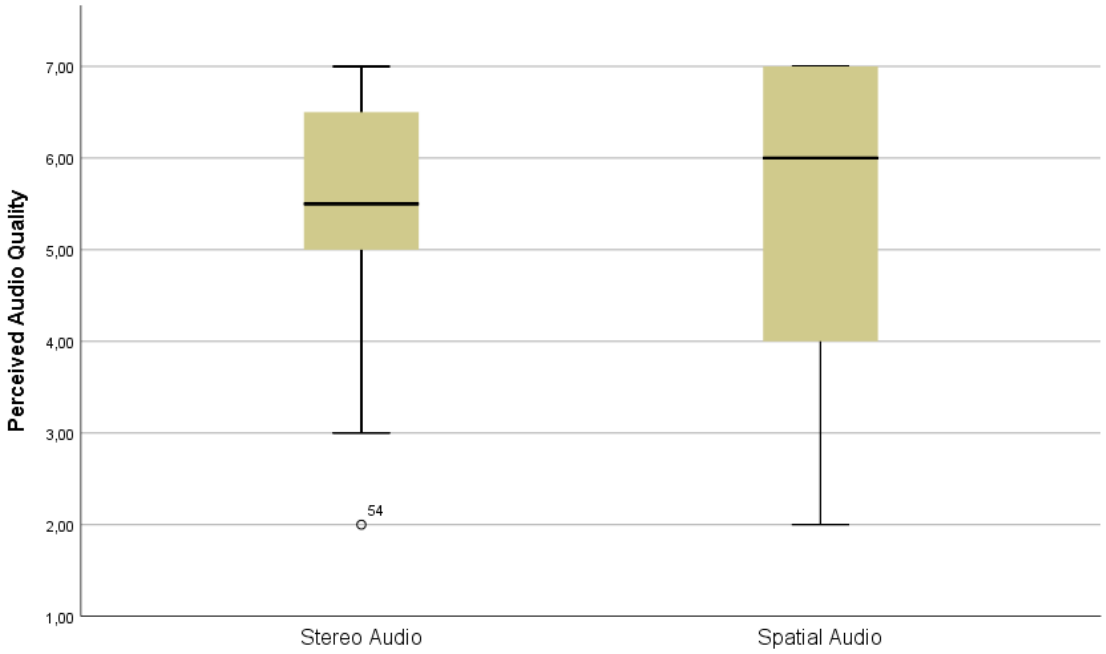

Figure 4.8 – Boxplot of perceived audio quality of audio formats paired with HMD

The largest difference between stereo and spatial audio comes in perceived stage presence where the minimum value recorded is four and 11 out 20 participants giving it the maximum value of seven compared to 8 out of 20 in stereo. While this outcome is expected – spatial audio in HMD scoring high in perceived presence- it remains a strong signal about spatial audio providing that extra kick for a live concert experienced in VR.

	Minimum	Maximum	Mean	Median
Stereo audio	3,00	7,00	5,92	6,00
Spatial audio	4,00	7,00	6,32	7,00

Table 4.11 – Perceived stage presence in audio formats paired with head mounted display (on the Likert scale)

The high scores of both audio formats are quite interesting. The distribution of the scores shown in figure 4.9 provides more information as spatial audio shows all participants giving a score of five or higher except for one outlier giving it the minimum score of four. Stereo audio has a more varied distribution of scores without any outliers present. The participant with the outlier score reported the audio coming from the *"wrong side"* compared to expectations possibly due to wearing the headset the other way around with the left speaker on the right ear and vice versa. This explains the lower score compared to the other participants. In spatial audio eleven participants (55%) gave a score of seven for perceived stage presence as compared to eight participants (40%) in stereo audio giving spatial audio an obvious advantage in perceived stage presence.

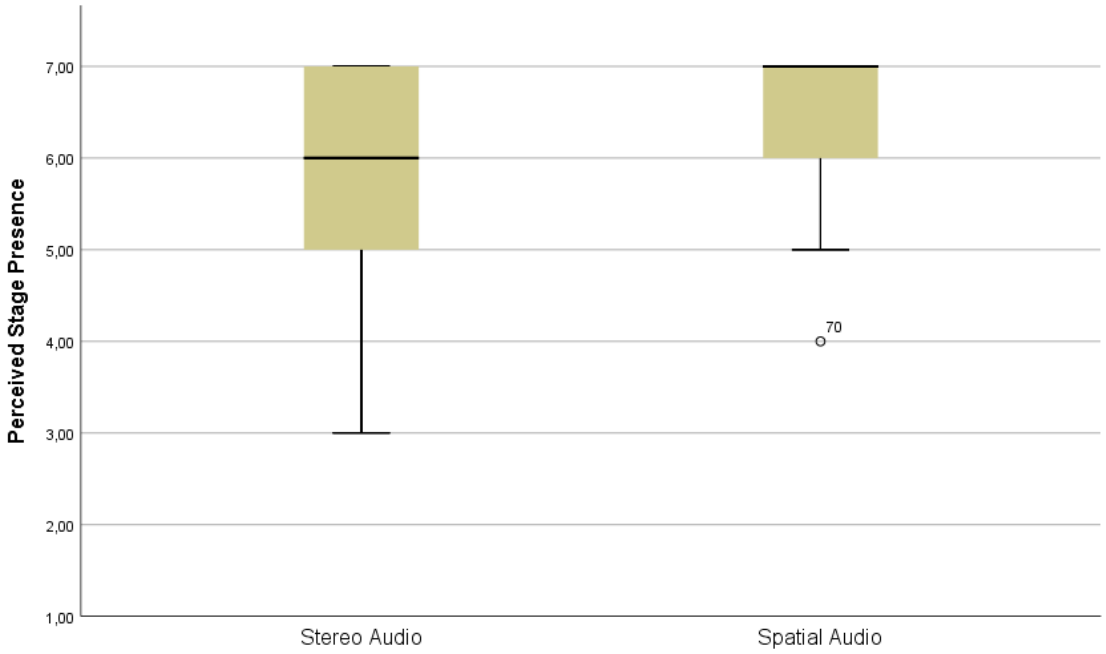

Figure 4.9 – Boxplot of perceived stage presence of audio formats paired with HMD

However the overall listening experience values reported indicate that the audio format did not have a large effect on the whole experience, as shown in table 4.16 with the means being so close they're almost negligible.

	Minimum	Maximum	Mean	Median
Stereo audio	3,00	7,00	5,80	6,00
Spatial audio	3,00	7,00	5,87	6,00

Table 4.12 – Overall listening experience in audio formats paired with head mounted display (on the Likert scale)

Despite the differences in other metrics between stereo audio and spatial audio with an HMD the overall listening experience seems to come down to similar results with similar score distribution. Thus giving both formats equal footing when it comes to an overall experience with HMD. The high means, medians, and score distribution of 75% participants giving a score of five or higher in both audio formats may be attributed to the HMD and its novelty "wow factor" effect.

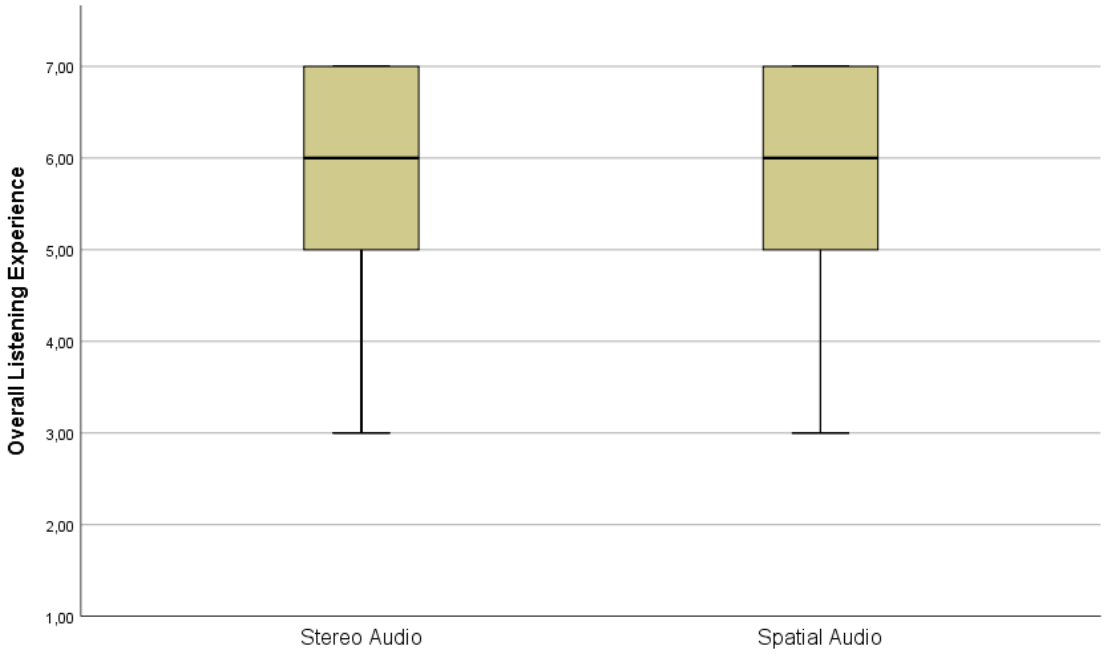

Figure 4.10 – Boxplot of overall listening experience of audio formats paired with HMD

The lack of a notable difference between stereo and spatial audio evaluations indicates that users may not care about spatial audio being available as it doesn't have such a significant impact on their whole experience while using a head mounted display. The figures below compare the means of the metrics from hobbyist and non-instrumentalist participants and that furthers doubt about the connection between instrument capabilities and favorable audio format, however as mentioned earlier in this document, this sort of connection is not a focus of this study.

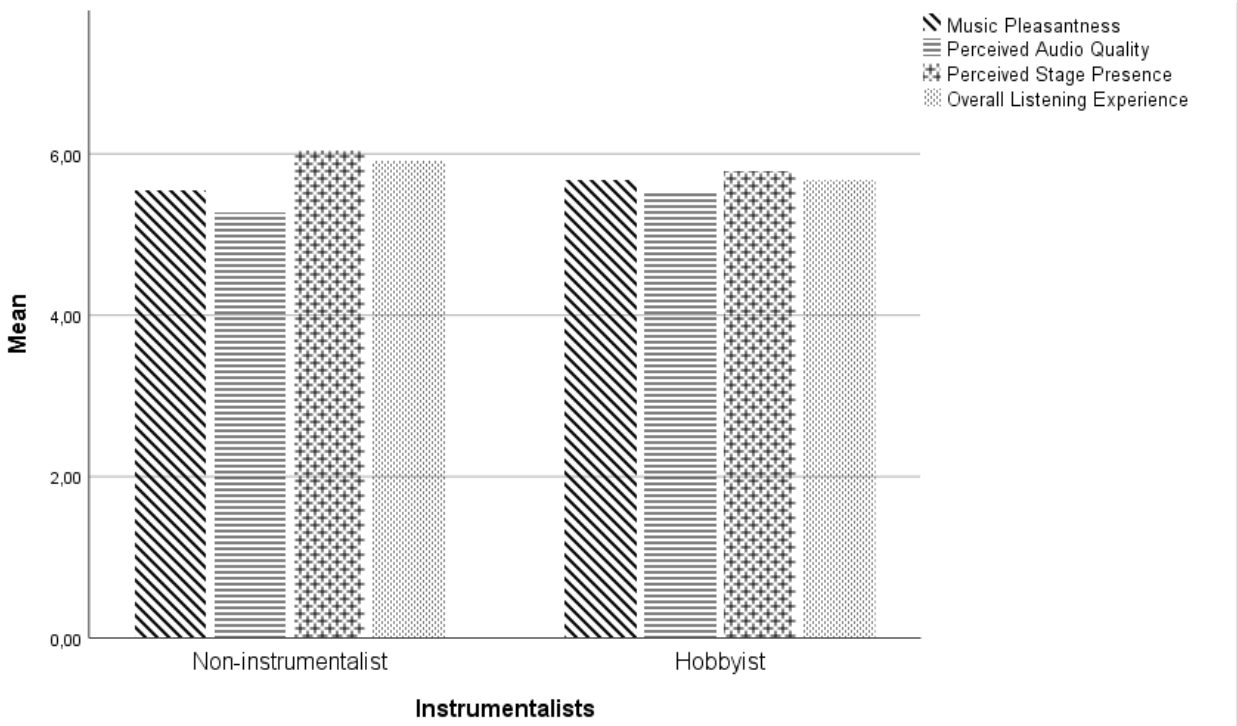

Figure 4.11 - Hobbyist and non-instrumentalist results means for stereo audio

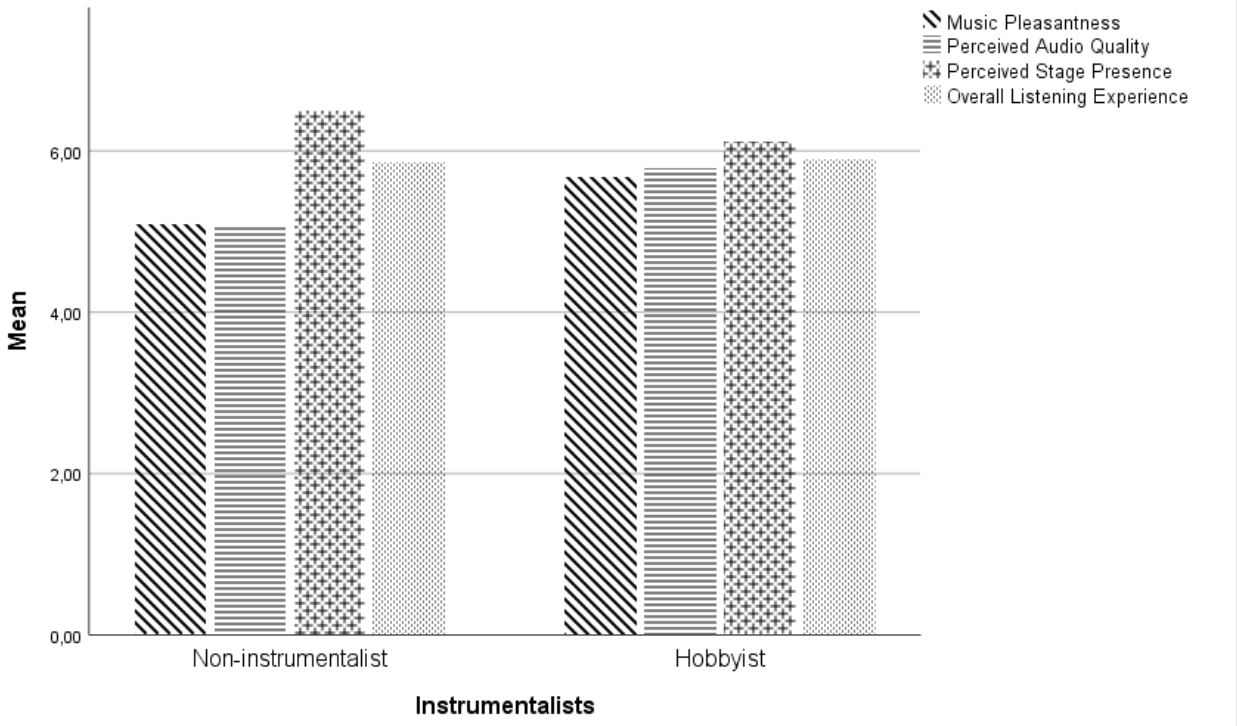

Figure 4.12 - Hobbyist and non-instrumentalist results means for spatial audio

4.3 Display comparisons (Flat display vs. HMD display)

This section compares the display performance and impact on the scores that it has paired with the two audio formats tested.

4.3.1 Stereo audio scenarios

An overall look at the mean differences of the metrics between flat and head mounted displays shows relatively dramatic differences in favor of head mounted display. The first of the metrics is music pleasantness, showing a difference of 1.2500 in favor of head mounted display. The impact may again be attributed to the novelty of VR as some of the participants had little to no experience with it at all.

	Minimum	Maximum	Mean	Median
Flat display	1,00	7,00	4,35	5,00
Head mounted display	2,00	7,00	5,60	6,00

Table 4.13 – Music pleasantness in displays paired with stereo audio (on the Likert scale)

The difference between the results is mainly in the distribution of the scores (shown in figure 4.13) as 75% of the participants gave a score of five or higher with HMD. While 75% of the participants with flat display gave a score below the median from the HMD results. All measures give HMD and advantage in music pleasantness in stereo audio.

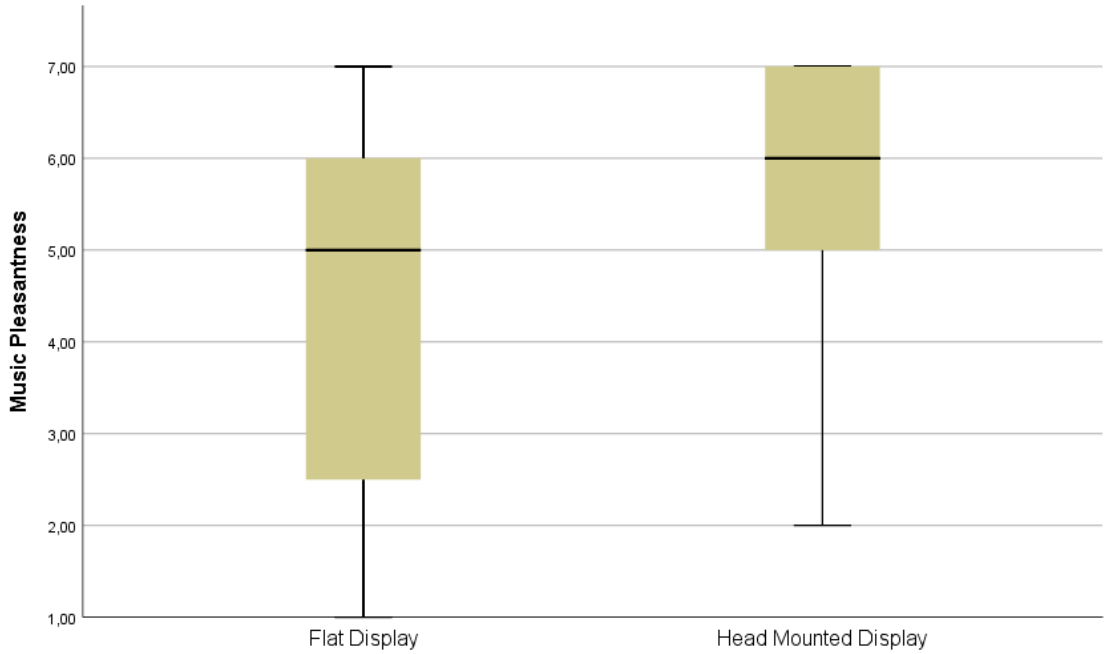

Figure 4.13 – Boxplot of music pleasantness in displays paired with stereo audio

The perceived audio quality is interesting because the video and the headset used remained the same, yet the perceived quality means had a jump of 0.7000 in favor of HMD. This is a further indicator that HMD is a stronger experience, however due to many participants being new to VR, the effects persisting over time and longer usage is an unknown measure.

	Minimum	Maximum	Mean	Median
Flat display	1,00	7,00	4,70	5,00
Head mounted display	2,00	7,00	5,40	5,50

Table 4.14 – Perceived audio quality in displays paired with stereo audio (on the Likert scale)

A further show of HMD's superior experience is the scores given as compared to flat display (figure 4.14). In HMD all participants but one outlier gave a score of three or higher with the majority of the participants giving a score of five or higher. Whereas in flat display there is a wider distribution with the scores ranging from one to seven. The majority of the participants gave a score of four or higher. Despite the difference in means between the two displays, and despite the impact the outlier has on HMD's mean in perceived audio quality, the medians are rather close with only a 0.5000 difference between the displays. The closeness of the measures alongside a marginal advantage for HMD may be related to one participant's comment regarding HMD giving the illusion of better audio quality. That may be attributed to the more immersive experience provided by HMD compared to flat display.

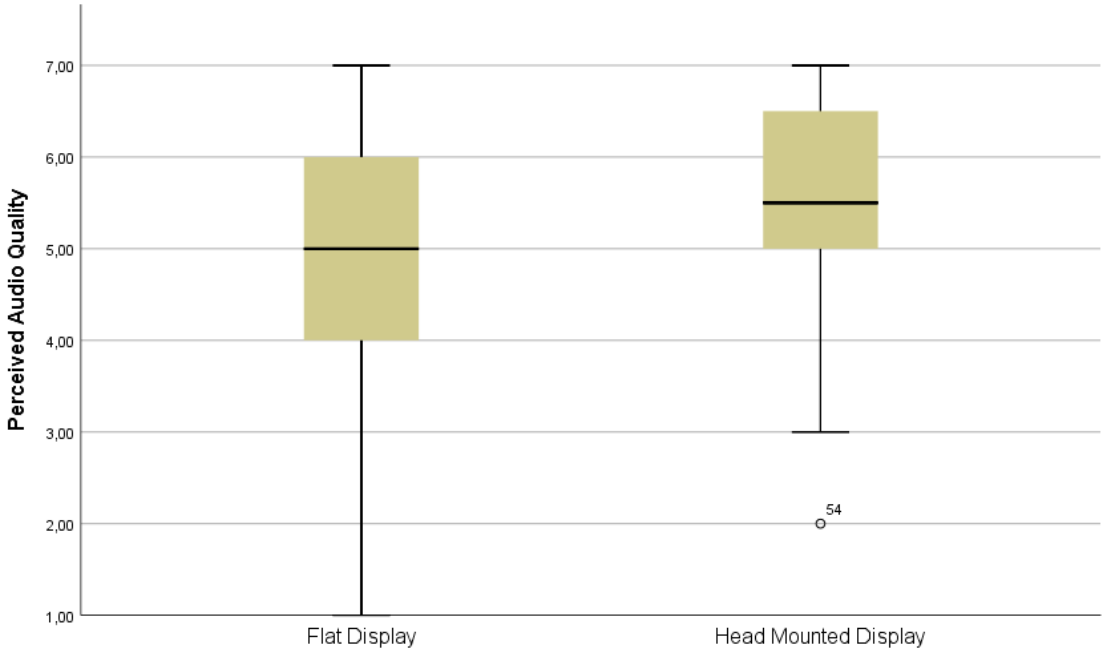

Figure 4.14 – Boxplot of perceived audio quality in displays paired with stereo audio

Table 4.19 shows rather unsurprising results with perceived stage presence as the nature of VR works towards a more immersive experience with a stronger sense of presence. The higher perceived presence from HMD allows for a more immersive experience encouraging people to acquire the hardware necessary to get the whole experience as such, one participant commented that they are *"going to think about buying a goggle"* referring to the Gear VR after the tests.

	Minimum	Maximum	Mean	Median
Flat display	1,00	7,00	3,82	4,00
Head mounted display	3,00	7,00	5,92	6,00

Table 4.15 – Perceived stage presence in displays paired with stereo audio (on the Likert scale)

As mentioned earlier, the large difference in perceived stage presence between the displays does not come as a surprise and the distribution of the scores (figure 4.15) provides further information regarding those differences. The responses are more varied towards flat display with 50% of the participants giving it a score between 2.5 and five with only 25% of the participants giving a score of five or higher, compared to 75% of the participants giving a score of five or higher in HMD, giving it a strong advantage for most participants.

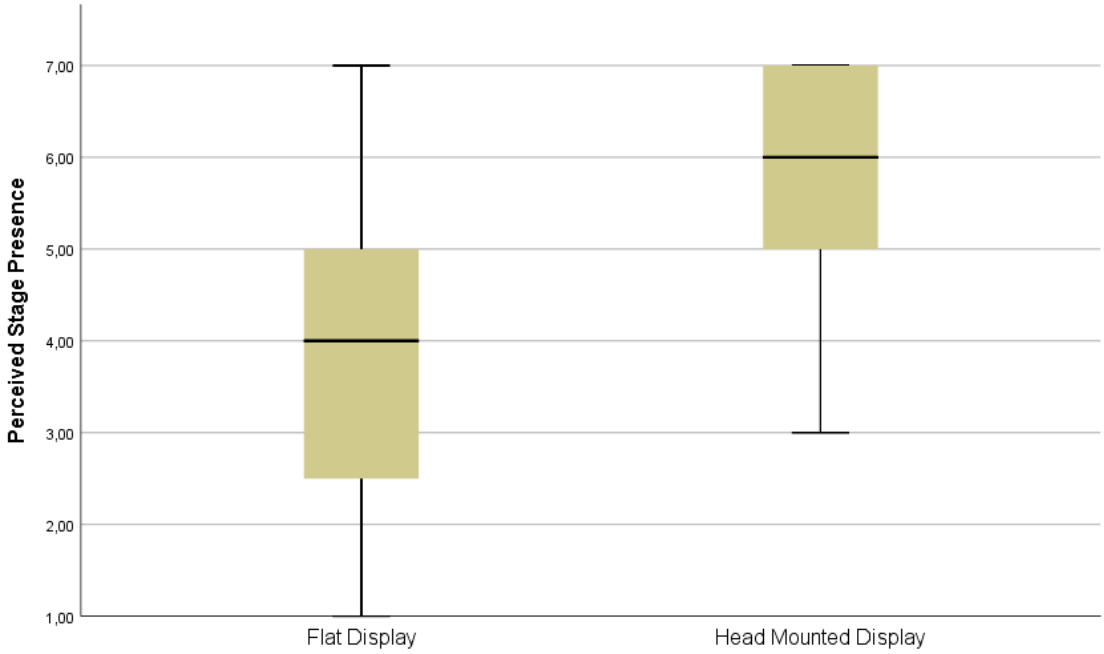

Figure 4.15 – Boxplot of perceived stage presence in displays paired with stereo audio

The same applies to the overall listening experience results shown in table 4.20, as HMD even allowed for some participants to move on from the fact they did not like or enjoy

the particular music selection of the test and were able to enjoy the experience as a whole without much regard of the content itself.

	Minimum	Maximum	Mean	Median
Flat display	1,00	7,00	4,20	4,50
Head mounted display	3,00	7,00	5,80	6,00

Table 4.16 – Overall listening experience in displays paired with stereo audio (on the Likert scale)

HMD's advantage from the previous metrics impact the result here as it maintains a strong one due to the closer range of responses from the participants compared to the more varied ones in flat display (as shown in figure 4.16).

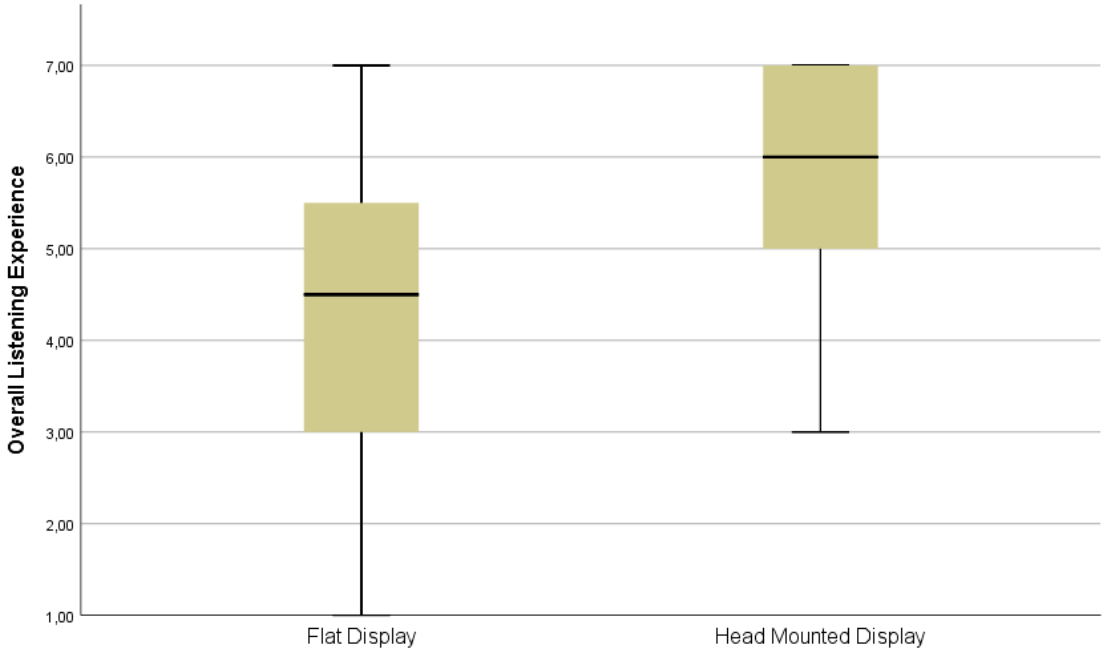

Figure 4.16 – Boxplot of overall listening experience in displays paired with stereo audio

4.3.2 Spatial audio scenarios

Overall HMD display outscores flat display in all metrics. Music pleasantness (shown in table 4.21) has a higher mean in HMD which may be attributed to the experience that HMD provides compared to the "normal" one from flat display.

	Minimum	Maximum	Mean	Median
Flat display	2,00	7,00	4,92	5,00

Head mounted display	2,00	7,00	5,35	6,00

Table 4.17 – Music pleasantness in displays paired with spatial audio (on the Likert scale)

The boxplot in figure 4.17 shows the closeness of the means between the displays is mostly due to two outliers giving a score of two for HMD thus impacting the mean of the results. The two outliers are the same participants discussed earlier as outliers, with one skipping the videos due to not liking the music choice at all. The other participant gave diplomatic responses in the post-interview which contradicted some of the scores given in the evaluation form such as this one. The rest of the participants gave a score of four or higher in HMD keeping the responses closer to each other compared to the participants' results spreading over a wider variation between the scores two and seven.

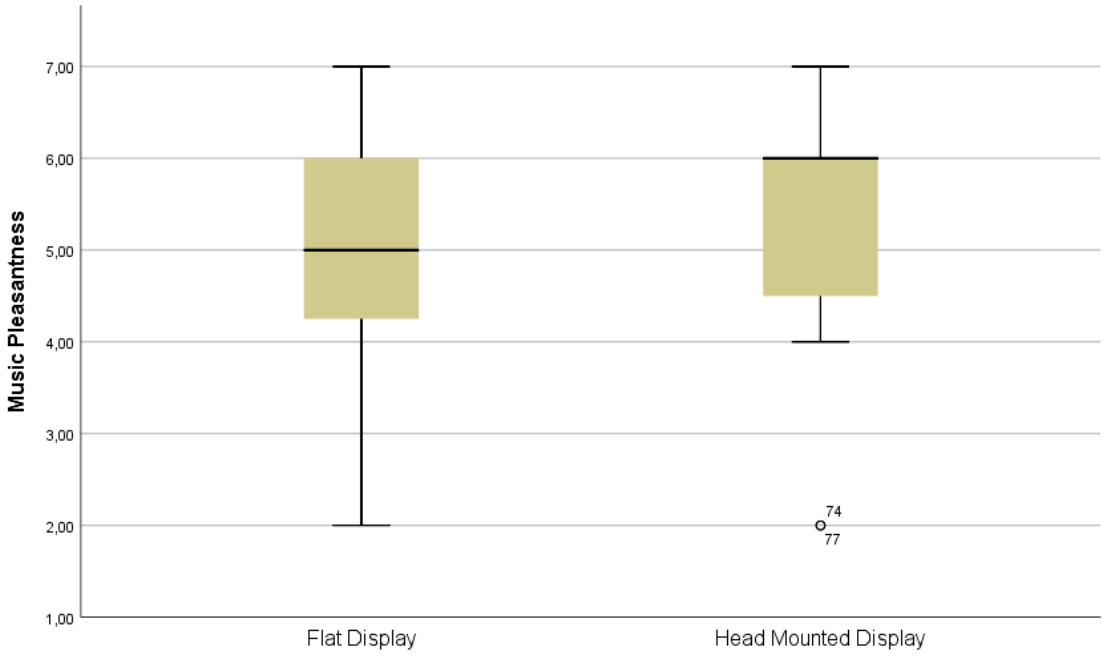

Figure 4.17 – Boxplot of music pleasantness in displays paired with spatial audio

Perceived audio quality results (presented in table 4.22) are interesting because despite the mean being higher in HMD, the minimum reported value is lower than that reported in flat display. The participant reporting that minimum score in HMD gave a value 4,00 for the same metric in flat display. Only one other participant perceived audio quality to be lower in HMD compared to flat display. These results may be due to the headset being worn the wrong way around as discussed earlier, which has a much higher impact with spatial audio as the sounds are programmed to come from the speakers accordingly.

	Minimum	Maximum	Mean	Median
Flat display	3,00	7,00	5,05	5,00

Head mounted display	2,00	7,00	5,40	6,00

Table 4.18 – Perceived audio quality in displays paired with spatial audio (on the Likert scale)

Despite the issues faced within the tests with some of the participants HMD still shows stronger results, which is a telling factor that having the right setup has a strong impact on the perceived results. The distribution of scores difference comes with HMD getting a score of 6 or higher with 50% of the participants compared to only 25% in flat display (shown in figure 4.18). And as the sound and audio quality are technically the same, this further shows that it may be true that HMD gives an illusion of better quality. However, whether the effects of that may persist with continuous use of the technology is not within the scope of this study, but is a curious topic to discuss in further studies.

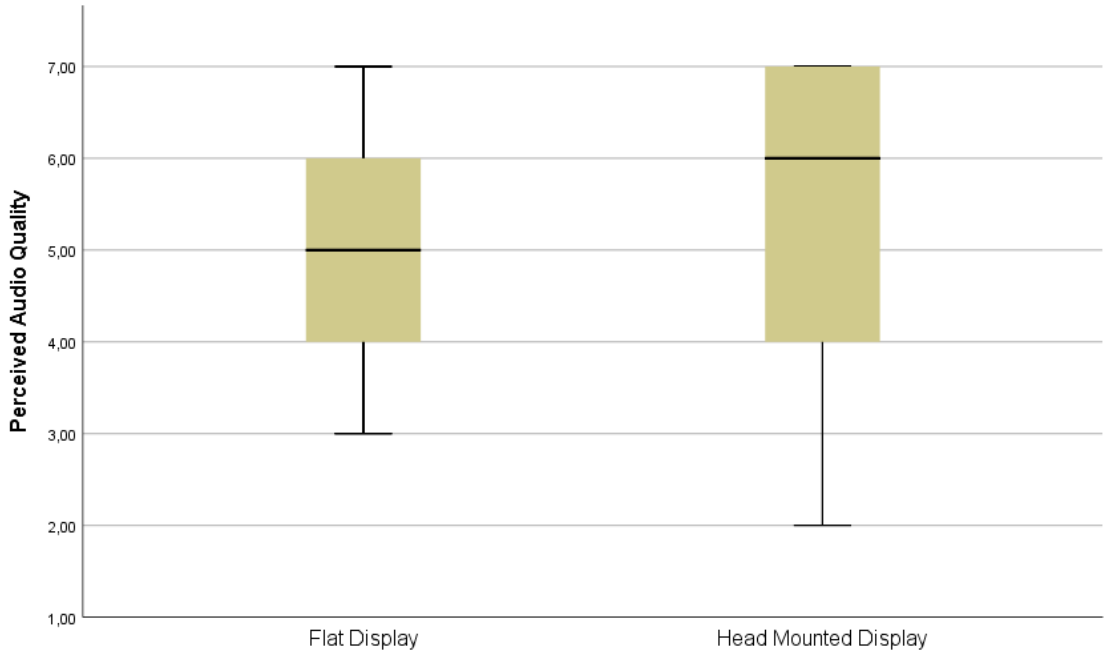

Figure 4.18 – Boxplot of perceived audio quality in displays paired with spatial audio

A great difference however with perceived stage presence (results shown in table 4.23) which does not come as a surprise as mentioned earlier on immersion using a head mounted display. Most notably the jump in minimum value recorded from one in flat display to four in HMD.

	Minimum	Maximum	Mean	Median
Flat display	1,00	7,00	4,29	5,00
Head mounted display	4,00	7,00	6,32	7,00

Table 4.19 – Perceived stage presence in displays paired with spatial audio (on the Likert scale)

The outlier in HMD is the same participant discussed earlier who reported the sound coming from the "wrong side" after watching the videos which is what possibly led to a break in perceived stage presence for them. The outlier and the distribution of the rest of the results in both displays are presented in figure 4.19 which shows the how strong the impact the display has on presence, with all the participants (except for the outlier) giving a score of five or higher in HMD, compared to widely varied and distributed responses in flat display ranging from one to seven.

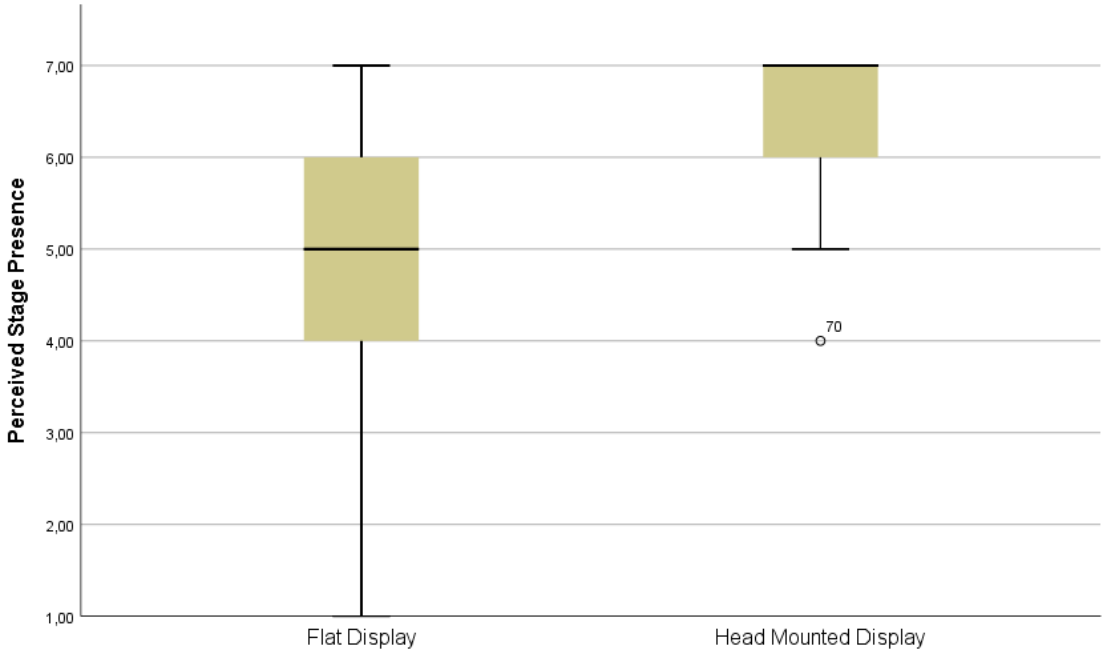

Figure 4.19 – Boxplot of perceived stage presence in displays paired with spatial audio

And finally the overall listening experience skewing towards HMD as shown in table 4.24.

	Minimum	Maximum	Mean	Median
Flat display	1,00	7,00	5,07	5,50
Head mounted display	3,00	7,00	5,87	6,00

Table 4.20 – Overall listening experience in displays paired with spatial audio (on the Likert scale)

More insight on the impact the display has on the overall experience is gathered from the boxplot in figure 4.20. Despite the medians being quite close, the distribution of scores tells a different story as 75% of the participants gave a score of five or higher to HMD

compared to 25% only for flat display. Thus giving a clear advantage for HMD towards a more widely better perceived overall experience to listen and/or to view a music video.

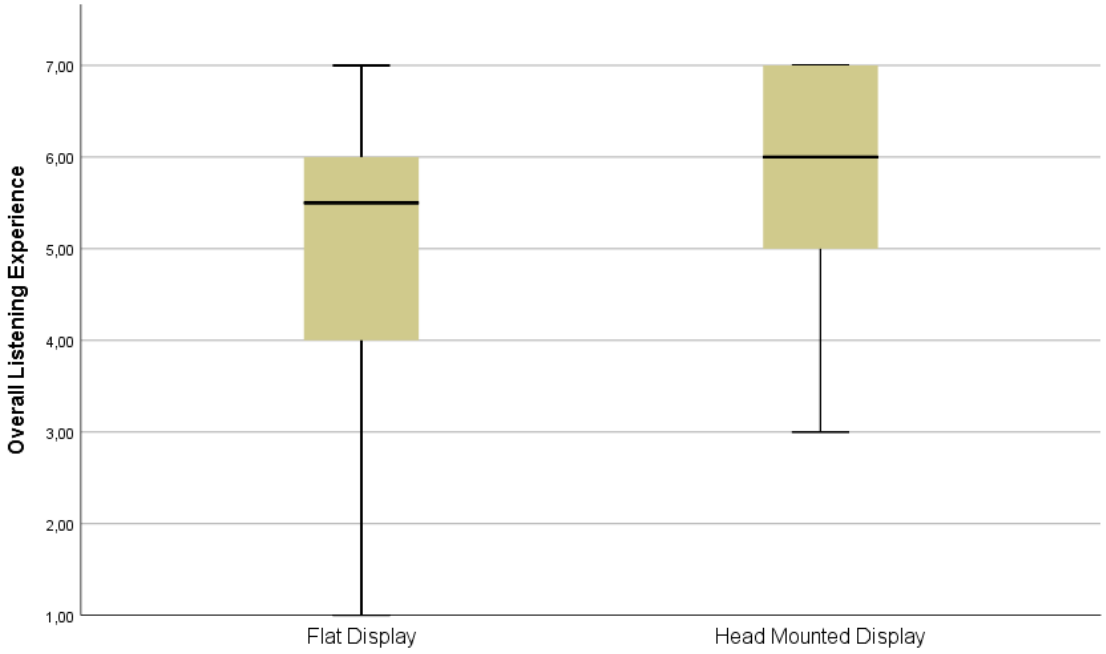

Figure 4.20 – Boxplot of overall listening experience in displays paired with spatial audio

4.4 Final Comparisons of the Scenarios

From the results of the four scenarios (display-audio format combinations) of each of the discussed metrics, the variations rank as follows for each metric. For **music pleasantness** the *HMD and stereo audio pairing has the highest mean* followed by HMD and spatial audio, flat display and spatial audio. Flat display and stereo audio scored the least mean value. The HMD and stereo audio pairing getting the highest music pleasantness mean is most curious however as discussed in earlier sections. The difference is due to two outliers in HMD and spatial audio results and the median is similar between the two audio formats with the display. The distribution of scores still gives stereo audio a marginal advantage over spatial audio and thus the *stereo audio still comes on top in this comparison with HMD* even with the outlies ignored.

Perceived audio quality *has both stereo and spatial audio formats paired with HMD as the variations with highest means* for this metric, followed by spatial audio in flat display and then stereo audio paired with the same display. This leads to possible confirmation of what one participant pointed out regarding HMD giving an illusion of better quality. In **perceived stage presence**, the educated guess that spatial audio and HMD would score highest is further confirmed, followed by stereo audio with HMD, and then spatial audio followed by stereo audio in flat display. *The ranking is similar for* **overall listening experience.**

Stereo audio paired with flat display scored the least in all metrics, and *spatial audio with HMD scored highest on all metrics apart from perceived audio quality*. That shows that *spatial audio with HMD would theoretically be the favored option* for an optimum experience out of the four options when watching a 360º music video, more specifically a live concert.

However after the participants finished all the variations and their evaluation forms, a **semi-structured interview** took place in order to get the personal preferences and further insight into what guided their answers, and what their explicit thoughts are on the different variations. *Three participants (15%) reported not hearing any difference at all between the different audio formats.* Some of the participants only noticed the differences after trying both variations of the audio. This gave a slightly clearer difference when testing with HMD. One participant commented that despite being unsure there was a difference in the listening experience from flat display, stereo audio stood out more and felt "boring" after listening to the music in spatial audio.

The results from the explicit preference and hindsight participant comparison cast some doubt on the evaluation results as they were not nearly as clear all across the board. The results are as follows, when asked about audio preference without regard to the display *15% (three participants) preferred stereo, 40% (eight participants) preferred spatial, and the remaining 45% (nine participants) were unspecified.* Unspecified means either the participant did not have a preference, or showed hesitation in making a decision without actually getting to one. *A preference does not dictate what the participant would want to listen to on a continuous basis*, as some of those who chose spatial as the preferred audio experience still thought stereo would be their preferred choice for a live concert.

When asked about the **display preference** without regard to the audio format, *18 out of the 20 participants preferred HMD VR*. The *novelty of the experience showed a large impact on the reactions of participants*. Further market research might support or debunk that. The result most telling and the one casting most doubt on the evaluation results conclusiveness is the one from asking the participant about the preferred experience from all four variations. From the participants, 55% were unspecified, five participants (25%) chose flat display and stereo audio. *Only two participants (10%) explicitly chose HMD and spatial audio as their preferred combination.* And so despite scoring the highest in overall experience from the evaluations, *it is not guaranteed as the choice way to experience listening to music* by most. The *number of participants preferring HMD and spatial audio increased to seven participants (40%)* when the question became more **specific about viewing live concerts** as opposed to general listening/viewing audiovisual experience purposes. The number of people preferring a flat display and stereo dropped to only one participant.

5. DISCUSSION & CONCLUSIONS

This is the final chapter of the document and it summarizes the findings from the study and the way the study was conducted, in addition to a discussion of the results and possible future work related to the subject.

5.1 Summary of Findings

A study was conducted with 20 participants to find the perceptions of differences in audio formats and displays across four different scenarios. The audio formats are spatial audio and stereo audio; and the displays are a flat display and head mounted display. Each scenario presents a combination of a display and an audio format. Perceptions of spatial audio are the main focus of this study, especially how those perceptions compare to the perceptions of stereo audio.

The final results show *HMD and spatial audio as the preferred combination with highest scores on all metrics*, except for **music pleasantness** in which HMD and stereo audio come on top with a very small margin. Results of the preferences interview show 11 participants (55%) did not specify a preference between any of the combinations. Only two participants (10%) chose HMD and spatial as their favorite. Five participants (25%) went with what they are used to (flat display and stereo audio) as their experience of choice.

5.2 Discussion

Throughout the scenarios and the different participants, three of the participants did not notice a difference between spatial and stereo audio. The participants reported that it sounded the same. Other participants reported that it was more apparent in VR due to easier movement and control over the environment when compared to using a mouse in flat display scenarios. The interaction with the flat display was referred to as *"unnatural"* by some of the participants.

The participants agreed that spatial audio would not be viable as a secondary task as it does not present the music piece completely *"as it is intended to be listened to"*. The quote gives way to the question of **"what would happen if spatial audio became the intended way to listen to music"**. The answer to the question requires testing with producers and sound engineers.

A large percentage of the participants use headsets for most of their music listening mostly due to a situation or environment that dictates the usage of headsets. That does not facilitate a switch to spatial audio. Listening to music is also widely viewed as a secondary

task, which would present the real area of switching habits for users when considering spatial audio as it works best as a main task.

Spatial audio does not work for a group of people to have a shared experience with the technology as it is now, especially with head mounted displays. As for a big screen the technology is getting there (e.g.: Dolby Atmos). The apparent lack of interest in audiovisual music listening experiences may be attributed to the seemingly dominant background listening behavior. In background listening, an audiovisual experience would interfere with the original main task at hand. Background listening as a habit appears to be key in other listening habits that participants showed and expressed. However we do not find the aforementioned alarming, especially when it comes to live concerts, as the value offered then is believed to be superior to the general lack of interest in a music video clip for example.

It is also important to take into account the "wow factor" for first time users of virtual reality. Spatial audio adds to the novelty of the experience which has an effect on the participants' responses. The long term reactions to the technology are not certain, as one participant described it as "*the illusion of better quality"* when it comes to VR. When asked about their most important take out from the experience, one participant commented that *"It was cool to experience VR glasses because I'd never used them before, and I had never seen 360° music videos so that was nice"* which further cements the observation of the impact of the novelty of the technology. However, the novelty impact from VR is stronger than that of spatial audio when results are compared separate of one another.

The pattern from the participants answers shows that spatial audio makes sense paired with the VR experience when compared to pairing it with a flat display. Participants also showed similar opinions when it came to a preference of audio format to use for their usual listening experiences, with the choice being stereo audio. However spatial audio is promising when it comes to live concerts due to its novelty and the higher presence and engagement it can help provide.

From the final preferences presented in the summary of findings subsection in this chapter, the reasoning behind not specifying a preference or choosing flat display with stereo audio wound down to at least one of the following reasons: 1) general doubt about the support of the technology and its availability. Participants 2) prefer the familiarity of stereo audio and flat displays. And finally 3) doubt about their own continued interest in the technology. The interest to experience live concerts in spatial audio and HMD is common between the participants as discussed earlier. Some participants expressed that they would like the choice to switch between the audio formats depending on their mood and preference at the time of experiencing it, while agreeing on HMD as the preferred display.

The results from this study regarding the perception of spatial audio especially when paired with VR are promising and encourage further studies of the topic. I believe that the right applications are key to the success of this technology in big markets and live entertainment is a recommended path to explore as based on the results of this study.

5.3 Limitations of the Study

The study presented multiple challenges and from each of those challenges a lesson was learned and will be helpful in future processes.

Theoretical background and information gathering – While the technologies had many online articles describing them, there was a notable lack of academic resources, especially when it came to applications and use cases (e.g.: VR in entertainment).

Interviews – Two semi-structured interviews are used in the tests in this study and while that provided good insight into the listening habits, the results from the post-test interview were hard to analyze and lead to doubts due to a lot of undefined answers. More structure to some of the questions in the interview could have provided more concrete answers.

The material and the participants – The material used in the test is in Finnish and many of the participants did not speak the language, a pre-requisite on the participants could have limited the variables – whether going for all non-speakers of the language or all native speakers of the language -.

The hardware – The head mounted display was on a buggy build that required some workarounds however that did not have an impact on the sessions. The room for user error with the headset placement in spatial audio lead to some outliers that would have been less likely to persist otherwise.

5.4 Conclusions and Future Work

The importance of listening habits comes in identifying the key aspects of understanding market for spatial audio. The listening habits can also provide towards a more user oriented spatial audio development. Spatial audio presents a new way to consume music and with targeting the right audience using the right applications, it can become a part of mass consumer markets.

Spatial audio is arguably not for background listening, thus it negates the possibility of listening to music in that form as a secondary task, necessitating dedicated time making the music listening the main task. A current percentage of dominantly dedicated listeners to music as from the test (10%) is a rather small niche when looking at the big picture of people listening to music all around.

With the rise of VR, and more concerts will be available to users to choose from in VR form, whether recorded or live. The availability allows for more people to choose spatial audio as it provides a stronger sense of presence and immersion with the concert they choose to “attend”. Availability also allows tapping into the 25% that are mood dependent when listening to music. It also allows for further switching of people who dominantly do dedicated listening, and from the majority of background listeners into mood dependent, or even dedicated listeners.

From the test, and from observing the participants, a plan to replace the current ways of listening would be nothing but a futile endeavor without much success. However, the technology complements the current ways of listening with an added depth to the experience. A wide variety and the availability of choices gives users the chance to experiment and adjust according to preference. It is recommended that spatial audio fills the gaps where stereo audio and background listening fall short, and to avoid attempting to replace stereo audio completely.

Another important conclusion to point out is the impact strength, the display impact seemed to be stronger than that of the audio format on the scores given to an experience. As mentioned earlier, that could be attributed to the novelty of VR and spatial audio. The novelty of VR had a much stronger impact on the participants than spatial audio. The difference in the impacts brings up a possibility that the impact of spatial audio is more easily sustainable, especially put together with VR for an experience catered for the users.

Suggested focuses for future spatial audio designs would be for bands near retirement, or bands and artist that no longer roam the earth, and those with fully sold-out concerts most of the time. Allowing the fans to be at the concert, from their home, within their convenience.

For an optimum experience users would ideally invest in better listening setups, leading more traffic into the Hi-Fi headphones market, and another segment of VR purchasing base that is driven by this format. Especially as spatial audio becomes more and more the go to choice to go with VR, rather than stereo audio.

Future research could focus on the artists and content creators and finding what value they can get out of spatial audio and how to encourage them to pursue more spatial audio production.

REFERENCES

[1] How did virtual reality begin? - Virtual Reality Society, *Virtual Reality Society*, 2017. [Online]. Available: https://www.vrs.org.uk/virtual-reality/beginning.html. [Accessed: 14-Feb-2019].

[2] Global augmented/virtual reality market size 2016-2022 | Statistic. [Online]. Available: https://www.statista.com/statistics/591181/global-augmented-virtual-reality-market-size/. [Accessed: 14-Feb-2019].

[3] Pimax: The World's First 8K VR Headset by Pimax 8K VR — Kickstarter. [Online]. Available: https://www.kickstarter.com/projects/pimax8kvr/pimax-the-worlds-first-8k-vr-headset. [Accessed: 11-Mar-2019].

[4] VR–1 – Varjo.com. [Online]. Available: https://varjo.com/vr-1/. [Accessed: 21-Mar-2019].

[5] Helvetin Pitkä Perjantai 360 Live - YouTube. [Online]. Available: https://www.youtube.com/watch?v=uuD4fPOhNog&t=35s. [Accessed: 06-Feb-2019].

[6] J. Holm and M. Malyshev, Spatial Audio Production for 360-Degree Live Music Videos, *2018 AES Int. Conf. Audio Virtual Augment. Real. (August 2018)*, pp. 1–3, 2018.

[7] J. Vince, *Introduction to Virtual Reality Concepts*. Springer, 2004.

[8] J. Jia and W. Chen, The ethical dilemmas of virtual reality application in entertainment, *Proc. - 2017 IEEE Int. Conf. Comput. Sci. Eng. IEEE/IFIP Int. Conf. Embed. Ubiquitous Comput. CSE EUC 2017*, vol. 1, pp. 696–699, 2017.

[9] M.-L. Ryan, *Narrative as Virtual Reality Immersion and Interactivity in Literature and Electronic Media*. The Johns Hopkins University Press, 2001.

[10] C. J. Fluke and D. G. Barnes, The Ultimate Display, *arXiv Prepr. arXiv1601.03459*, pp. 1–2, 2016.

[11] F. P. Brooks, What's real about virtual reality?, *IEEE Comput. Graph. Appl.*, vol. 19, no. 6, pp. 16–27, 1999.

[12] D. Grigorovici, Affectively engaged: Affect and arousal routes of entertainment virtual reality, *Proc. - 7th Int. Conf. Virtual Syst. Multimedia, VSMM 2001*, pp. 634–643, 2001.

[13] J. Steuer, Defining virtual reality: dimensions determining telepresence, Communication in the age of virtual reality, *J. Commun.*, vol. 42, no. 4, pp. 73–93, 1992.

[14] M. Slater, A note on presence, *Presence Connect*, vol. 3, pp. 1–5.

[15] New partnership brings virtual reality to NBA on TNT | NBA.com. [Online]. Available: http://www.nba.com/article/2017/11/07/new-partnership-brings-virtual-reality-nba-tnt. [Accessed: 09-Oct-2018].

[16] 360 Labs | a VR Production Company. [Online]. Available: https://360labs.net/. [Accessed: 11-Mar-2019].

[17] Filmed entertainment revenue worldwide 2015 | Statistic. [Online]. Available: https://www.statista.com/statistics/259985/global-filmed-entertainment-revenue/. [Accessed: 09-Oct-2018].

[18] Games industry generated $108.4bn in revenues in 2017 | GamesIndustry.biz. [Online]. Available: https://www.gamesindustry.biz/articles/2018-01-31-games-industry-generated-usd108-4bn-in-revenues-in-2017. [Accessed: 09-Oct-2018].

[19] Global music industry grew 8% last year jazzed by streaming subscriptions | Financial Times. [Online]. Available: https://www.ft.com/content/6f395204-47ca-11e8-8ee8-cae73aab7ccb. [Accessed: 09-Oct-2018].
[20] D. Dolan and M. Parets, Redefining The Axiom Of Story:The VR And 360 Video Complex, 2016. [Online]. Available: http://techcrunch.com/2016/01/14/redefining-the-axiom-of-story-the-vr-and-360-video-complex/.
[21] VR Fact Sheet 2018-An Overview of VR Films, Games Experiences, in *2018 IEEE Games, Entertainment, Media Conference, GEM 2018*, 2018, pp. 122–124.
[22] Ways VR is Changing the Music Industry - Mbryonic. [Online]. Available: https://mbryonic.com/music-vr/. [Accessed: 06-Feb-2019].
[23] Gorillaz - Saturnz Barz (Spirit House) 360 - YouTube, 2017. [Online]. Available: https://www.youtube.com/watch?v=lVaBvyzuypw. [Accessed: 06-Feb-2019].
[24] Metallica: Sad But True (360° Video) (Stockholm, Sweden - May 7, 2018) - YouTube. [Online]. Available: https://www.youtube.com/watch?v=yUpex6N9qyk. [Accessed: 06-Feb-2019].
[25] Metallica: Seek & Destroy 360° (Foxborough, MA - May 19, 2017) - YouTube. [Online]. Available: https://www.youtube.com/watch?v=Q2WhXnVM0f4. [Accessed: 06-Feb-2019].
[26] Megadeth - Poisonous Shadows (Live) (360) - YouTube. [Online]. Available: https://www.youtube.com/watch?v=unQZvhQ9Giw. [Accessed: 06-Feb-2019].
[27] Sean G. Kelly. [Online]. Available: https://seangarrettkelly.com/#/teachuvr/. [Accessed: 07-Feb-2019].
[28] Save 50% on Electronauts on Steam. [Online]. Available: https://store.steampowered.com/app/691160/Electronauts/. [Accessed: 07-Feb-2019].
[29] U. Shukla, An introduction to 360° video | Knight Lab Studio. [Online]. Available: https://studio.knightlab.com/results/storytelling-layers-on-360-video/an-introduction-to-360-video/. [Accessed: 11-Mar-2019].
[30] Professional 360 Camera Guides - Find the right camera for your 360 project - 360° Camera Reviews and Guides. [Online]. Available: http://www.threesixtycameras.com/professional-360-cameras/. [Accessed: 11-Mar-2019].
[31] MagentaMusik 360 | Telekom. [Online]. Available: https://www.magenta-musik-360.de/. [Accessed: 11-Mar-2019].
[32] Pure McCartney VR - virtual reality video | Jaunt. [Online]. Available: https://www.jauntvr.com/lobby/PaulMcCartney. [Accessed: 11-Mar-2019].
[33] VR film company Jaunt is giving up on VR to focus on augmented reality - The Verge. [Online]. Available: https://www.theverge.com/2018/10/15/17980420/jaunt-vr-layoffs-ar-focus-switch-restructuring-xr-platform. [Accessed: 11-Mar-2019].
[34] J. Habig, Is 360 Video Worth It?, *think with Google*, no. June, 2016.
[35] M. Hosseini and V. Swaminathan, Adaptive 360 VR video streaming: Divide and conquer!, *Proc. - 2016 IEEE Int. Symp. Multimedia, ISM 2016*, pp. 107–110, 2017.
[36] Introducing Spatial Audio for 360 Videos on Facebook. [Online]. Available: https://www.facebook.com/facebookmedia/blog/introducing-spatial-audio-for-360-videos-on-facebook. [Accessed: 25-Jan-2019].
[37] Sound Systems: Mono vs. Stereo. [Online]. Available: http://www.mcsquared.com/mono-stereo.htm. [Accessed: 25-Jan-2019].
[38] 5.1 vs. 7.1 Home Theater Explained | The Master Switch. [Online]. Available:

https://www.themasterswitch.com/51-vs-71-home-theater-explained. [Accessed: 11-Mar-2019].

[39] What is Dolby Atmos? All you need to know | Trusted Reviews. [Online]. Available: https://www.trustedreviews.com/opinion/dolby-atmos-2942509. [Accessed: 11-Mar-2019].

[40] Ambisonics Explained: A Guide for Sound Engineers | Waves. [Online]. Available: https://www.waves.com/ambisonics-explained-guide-for-sound-engineers. [Accessed: 26-Jan-2019].

[41] An Introduction to Ambisonics | Creative Field Recording. [Online]. Available: https://www.creativefieldrecording.com/2017/03/01/explorers-of-ambisonics-introduction/. [Accessed: 26-Jan-2019].

[42] Binaural audio: What is it? How can you get it? | What Hi-Fi? [Online]. Available: https://www.whathifi.com/advice/binaural-audio-what-it-how-can-you-get-it. [Accessed: 27-Jan-2019].

[43] A. Farina, Introducing SPS format and its first practical implementation, called Mach1. [Online]. Available: http://pcfarina.eng.unipr.it/SPS-conversion.htm. [Accessed: 29-Jan-2019].

[44] Get OZO Audio, licensable spatial audio technology for any recording device | Nokia OZO. [Online]. Available: https://ozo.nokia.com/en/products/ozo-audio.html?gclid=EAIaIQobChMI9Lae8J-M4AIVSKqaCh1NVAq0EAAYASAAEgJygPD_BwE. [Accessed: 29-Jan-2019].

[45] E. L. Tan, W. S. Gan, and C. H. Chen, Spatial sound reproduction using conventional and parametric loudspeakers, *Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC), 2012 Asia-Pacific*, pp. 1–9, 2012.

[46] M. J. Morrell and J. Reiss, Ambisonics based Music Composition and Production Tool for an Octagonal Speaker Layout, *C. 2012 Music Emot. 2012*, no. June, pp. 233–240, 2012.

[47] HTC Vive Introduce Spatial Audio SDK – VRFocus. [Online]. Available: https://www.vrfocus.com/2018/06/htc-vive-introduce-spatial-audio-sdk/. [Accessed: 25-Jan-2019].

[48] Spatial Audio API | Google VR | Google Developers. [Online]. Available: https://developers.google.com/vr/reference/ios-ndk/group/audio. [Accessed: 25-Jan-2019].

[49] A. Tse, C. Jennett, J. Moore, Z. Watson, J. Rigby, and A. L. Cox, Was I There?, in *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '17*, 2017, pp. 2967–2974.

[50] R. B. Lind *et al.*, Sound design in virtual reality concert experiences using a wave field synthesis approach, in *Proceedings - IEEE Virtual Reality*, 2017, pp. 363–364.

[51] R. L. Storms and M. J. Zyda, Interactions in perceived quality of auditory-visual displays, *Presence Teleoperators Virtual Environ.*, vol. 9, no. 6, pp. 557–580, 2000.

[52] J. H. Chang and W. H. Cho, Impairments of binaural sound based on ambisonics for virtual reality audio, *Proc. IEEE Sens. Array Multichannel Signal Process. Work.*, vol. 2018–July, no. Ild, pp. 341–345, 2018.

[53] Back in the groove: How vinyl rose from its sickbed to capture the eyes and ears of millennials | The Independent. [Online]. Available: https://www.independent.co.uk/news/long_reads/vinyl-demand-lps-record-store-day-a7952911.html. [Accessed: 19-Nov-2018].

APPENDIX A: BACKGROUND QUESTIONNAIRE

1. Age: _____

Background Information Form

2. Gender:
   - ❍ Male
   - ❍ Female

3. Education:
   - ❍ Comprehensive or elementary school
   - ❍ High School
   - ❍ College/University Degree
   - ❍ Other: _____

4. How familiar are you with Virtual Reality devices:

	1	2	3	4	5	6	7	
Not Familiar	❍	❍	❍	❍	❍	❍	❍	Familiar

5. How familiar are you with 360° videos:

	1	2	3	4	5	6	7	
Not Familiar	❍	❍	❍	❍	❍	❍	❍	Familiar

6. How familiar are you with 360° music videos:

	1	2	3	4	5	6	7	
Not Familiar	❍	❍	❍	❍	❍	❍	❍	Familiar

7. How familiar are you with spatial audio:

	1	2	3	4	5	6	7	
Not Familiar	❍	❍	❍	❍	❍	❍	❍	Familiar

8. Do you play any musical instruments?

❍ No
❍ Hobbyist player
❍ Professional musician

APPENDIX B: VIDEO EVALUATION FORM

*DO NOT FILL
Participant number: _____

❍ PC - Stereo
❍ PC - Spatial
❍ VR - Stereo
❍ VR - Spatial

1. How pleasant was the music?

	1	2	3	4	5	6	7	
Very unpleasant	❍	❍	❍	❍	❍	❍	❍	Very pleasant

2. How good was the audio quality?

	1	2	3	4	5	6	7	
Very poor	❍	❍	❍	❍	❍	❍	❍	Very good

3. How present did you feel amongst the musicians on stage?

	1	2	3	4	5	6	7	
Not present at all	❍	❍	❍	❍	❍	❍	❍	Completely present

4. How pleasant was the overall listening experience with the video?

	1	2	3	4	5	6	7	
Very unpleasant	❍	❍	❍	❍	❍	❍	❍	Very pleasant

5. How did the choice of music affect the overall experience?

	1	2	3	4	5	6	7	
Negatively	❍	❍	❍	❍	❍	❍	❍	Positively

APPENDIX C: INTERVIEW QUESTIONS BEFORE TEST COMMENCING

- When do you listen to music? (e.g.: Studying, Sports, Dedicated listening, Bus, etc.)
- How do you listen to music? (Headset vs. Speakers) (Alternatively: What's your music setup?)
- How do you listen to music? (Background vs. dedicated)
- How do you listen to music? (Audio vs. with video)
- How does the quality/resolution of the music affect your listening experience? (to detect any hi-fi people)
- What do you know about spatial audio? (if any)

APPENDIX D: INTERVIEW QUESTIONS

- Which one is the preferred one (computer vs. goggles)
- How did you feel about Spatial audio?
- How does it feel compared to stereo audio?
- Which one would you choose and why (spatial vs stereo)?
- How do you see yourself using spatial audio?
- Would you be interested in seeing whole 360 concerts with spatial audio? How about stereo? (Does it really matter which audio experience it is)
- What are the advantages you see for spatial audio over stereo?
- Do you see yourself switching to spatial audio in your listening time? Why (not)?
- What was your most important take out from the experience?
- What music or concerts would you like to hear in Spatial audio
- Which instruments sounded like they were moving for you (and how pleasant or unpleasant was it)